Research universities are viewed as one of the primary drivers of the knowledge economy worldwide. Based on the Humboldtian idea of combining research and teaching, along with a high degree of autonomy and academic freedom, these universities are the flagship institutions in any national higher education system. There is strong evidence from many countries, both developing and developed, that a period of establishing new research universities and strengthening existing ones has begun (Altbach and Salmi 2011; Liu, Wang, and Cheng 2011). Moreover, the future of the research university as a model is considered to be bright and promising despite the tremendous changes in the field of higher education. Altbach (2011) notes that it is unlikely that research universities will be fundamentally different in 2050 than they are today. What factors contribute to the development of an excellent research university? Obviously, there is no universal recipe for the creation or development of such institutions, although they have commonalities. Salmi (2009) identifies three sets of features that distinguish world-class research universities. The first feature is the high concentration of talent among both faculty and students. Leading research universities can attract the best students (with an emphasis on postgraduates) and most-qualified professors and researchers at the national and international levels. The second feature is the abundant resources that are necessary for the creation of a suitable environment for cutting-edge research and the generation of new knowledge. The third feature is favorable governance, which is apparent in the high degree of academic and managerial autonomy of research universities. This type of governance should encourage academic freedom, flexibility, organizational learning and a culture of excellence. The development of a research university (particularly one that aspires to be a world-class research university) is a challenging task for both the national government and university leadership. There can be many bumps on the ‘road to academic excellence’. How does the institution attract the best students and faculty? How can the institution obtain adequate and stable funding? How does the institution address problems of privatization and globalization (Altbach 2011)? Research universities are becoming increasingly complex organizations embedded in knowledge networks. Facilitating change and improvement in such networks implies strategic decisions with high stakes. The present article elaborates on one of the basic elements of the decision making process of any leading research university, namely, the practice of institutional research (IR). In essence, IR ‘has to do with what decision makers need to know about an institution, its educational objectives, goals and purposes, environmental factors, processes and structures to more wisely use its resources, more successfully attain its objectives and goals, and to demonstrate integrity and accountability in so doing’ (Dressel 1981, 237). By definition, IR is closely connected to the above-described three sets of features that characterize a world-class research university. IR may be useful in informing decisions regarding the attraction and retention of ‘the best and brightest’, fundraising, resource use, and the development of a favorable governance environment. This article addresses three questions. First, how did IR arise, and what is its raison d’être in a research university? Second, how can IR contribute to the improvement of the research university? Finally, what are the most viable alternatives with regard to the structure, staffing, and responsibilities of the IR office? To answer these questions, the article will draw on the historical and current state of IR data from different countries derived from an extensive literature review and several case studies conducted while launching and running the IR office in a newly established research university, the Higher School of Economics (the HSE) in Moscow, Russia. In a relatively short period of time, this university became one of the largest research universities in the field of social and economic sciences in Eastern Europe (Froumin 2011). Additionally, the practice of IR played an important role in the strategic development of the HSE. Therefore, the findings obtained from this study may be relevant to research universities in developing and middle-income countries. The origins of and rationale for IR What is IR? How old is IR? Why has IR been introduced in hundreds of universities globally? Although there are as many possible answers to these questions as there are institutional researchers, in this section, I will summarize the primary points of discussion regarding the characteristics and background of IR from an international perspective. Defining IR is not an easy task. Most definitions emphasize the functional aspects of IR and note that it is directed towards the support of management at an institution: ‘what distinguishes institutional research stricto sensu is its application to the individual establishment of higher education’ (Neave 2003, 3–4). Saupe (1990) provides a similar and highly influential definition of IR: ‘research conducted within an institution of higher education in order to provide information which supports institutional planning, policy formulation and decision making’ (1). There have also been efforts to provide a generalized definition that is broad enough to accommodate various functions and characteristics of IR. Terenzini (1993) considers IR to be ‘organizational intelligence’ and states that it refers ‘to the data gathered about an institution, to their analysis and transformation into information, and to the insight and informed sense of the organization that a competent institutional researcher brings to the interpretation of that information’ (3). Recent studies demonstrate that the activities attributed to IR evolve over time (Peterson 1999; Lasher 2011). Moreover, there is a significant variation among universities and national systems of higher education in definitions and functions of IR (Peterson 2003; Ferren and Merrill 2012). The nature of IR greatly depends on the environment in which the institution is embedded and is affected by the decisions of federal governments, student characteristics, changes in technology, decision-making processes, the location of the IR office in the university structure, and other factors (Neave 2003; Volkwein, Liu, and Woodell 2012). Therefore, IR is a contextual phenomenon and must be explored in particular settings. Further, the article provides an overview of the emergence and primary goals of IR in various regions of the world. It is widely recognized that US universities are pioneers in the development of IR as a formally defined set of functions within an institution (Reichard 2012). Some scholars date the appearance of sporadic IR projects in American universities to the eighteenth century (Cowley 1960; Saupe 2005). Thus, the history of such research spans over three centuries. However, the process of recognizing IR as an independent field of practice began in the USA a relatively short time ago, during the 1920s. At that time, Bureaus of IR were organized at several universities. The 1960s are described as a period of significant growth for IR, when several national conventions, regional meetings, and workshops occurred and the Association for Institutional Research (AIR) was established (Reichard 2012). The role of IR was initially limited to producing descriptive statistics for university leadership and fact books for various stakeholders. Institutional researchers then began to engage more in analysis and evaluation, strategic planning, and academic program reviews. Currently, the function of IR in US universities is threefold: (1) institutional reporting and administrative policy analysis; (2) strategic planning, enrolment, and financial management; and (3) outcome assessments, program reviews, accountability, accreditation, and institutional effectiveness (Volkwein, Liu, and Woodell 2012). In Western European countries, the development of IR was slightly delayed and began during the second half of the twentieth century. IR in the UK and Sweden evolved out of various ‘committees of enquiry’ established by governmental bodies to address the expanding demand for higher education (Neave 1989, 211–22). In France, Italy, and Spain, the interest in IR began two decades ago and was spurred by a decentralization of the higher education system that was aimed towards solidifying the connection between universities and local government agencies/communities. This decentralization led to the appearance of new interested parties: stakeholders that demanded transparency and reports on university efficiency (Neave 2003). IR in Europe is different from that in the USA and is less scholarly and more focused on policy and management issues; in many cases, IR is not a separate function within the university administration but the collective responsibility of university managers. If we compare the membership structure of the AIR and its European counterpart, the European Association of Institutional Research (EAIR, founded in 1979), the latter is more an association of administrators and governmental policy analysts, whereas AIR includes many junior analysts and institutional researchers (Peterson 2003). Universities in Australia and New Zealand also demonstrate a strong interest in developing IR capacity. In Australia, IR began to develop during the early 1970s, when the Australian government required greater accountability from the institutions in the form of student, staffing, and financial statistical information (Maasen and Sharma 1991). Currently, many universities have separate IR units that address a wide range of tasks, including the coordination of strategic planning, quality assurance, managing student life-cycle surveys, statistical analysis for internal and external reporting, resource allocation and planning, and other areas (Sharma 2010). In New Zealand, a deregulated student demand-driven funding framework initially introduced in the 1990s, followed by reforms focused on performance incentives and accountability, became important factors that facilitated the development of IR (Hanlon, Rothery, and Daldy 2011). IR professionals in the Australasian region cooperate and raise their level of competence through participation in the Australasian Association for Institutional Research (AAIR), which was established in 1988. Despite the tardy arrival of other countries and regions to the field of IR, the last decade has seen a tremendous global growth in the number of universities eager to broaden IR agendas to inform their planning and decision making. One indicator of the professionalization of IR is the emergence of several professional associations, which are active in promoting the culture of research-driven decision making. The most visible organizations of this type are the Southern African Association for Institutional Research (founded in 1994), the Canadian Institutional Research and Planning Association (founded in 1994), the Higher Education Research and Policy Network based in Nigeria (founded in 2000), the South East Asian Association for Institutional Research (founded in 2001), and the recently established (2009) Middle East and North Africa Association for Institutional Research (MENA-AIR). All of these organizations are affiliated with AIR and promote IR practice by organizing annual conferences, seminars, and workshops and publishing newsletters and academic journals. The main characteristic of the development of IR in these regions is that universities in many respects adopt the best practices in IR from US and European universities, enabling the professionalization of IR to occur more rapidly. For example, a recent study conducted by MENA-AIR showed that although approximately half of the IR offices in Middle Eastern and North African countries have been established in the last five years, the characteristics and responsibilities of these IR offices are similar to those in the USA (Volkwein, Liu, and Woodell 2012). Simultaneously, the IR agendas of these offices often reflect national higher education trends and policies. A basic level of IR can be found in many other countries that are not actively involved in any national or international professional associations. Latin American universities demonstrate some activities that can be attributed to IR, although the level of professionalization is generally low (Carranza, Durland, and Corengia 2010). Several Chinese universities began practicing IR as early as the 1990s but remain in the initial stages of development (Chang 2009). In Russia, IR is also a relatively new activity and has thus far been implemented as a specific function in a small number of leading state research universities and several private institutions based on the US or UK model of higher education. Few Russian universities have a separate unit for IR; the primary responsibilities are associated with collecting data for external reporting and accreditation. In the HSE, the IR office was established in 2009 and focuses on statistical analyses, sociological surveys, and qualitative studies for university administration decision-making processes, which are generally unusual practices for Russian universities. Originally, these practices were an initiative of a group of sociology students who began conducting student and faculty surveys for university improvements in 2001. Subsequently, when the university had observed a tremendous growth in the number of students, faculty, and educational programs and was awarded a special status, the National Research University (with more funding and autonomy, but also more accountability), university leadership decided to develop an IR office based on this group of recent graduates. This brief overview of the development of IR in different countries and regions shows that it is truly a global phenomenon. Therefore, despite some variations at the institutional and national levels, there must be a number of global driving forces in higher education that facilitate IR. Research universities that serve as nodes of the international intellectual community are primarily exposed to these driving forces. It is proposed that two sets of factors emphasize the process of IR emergence and development in research universities. The first set includes factors that facilitate IR in general at all types of higher education institutions, such as massification, the marketization of higher education, and the changing nature of the relationships between universities and governments. The second set includes factors that are more specific to research universities: increasing internationalization and the emergence of global rankings. However, some factors may be more salient than others in specific cases of institutions, and the internal characteristics of a university (e.g. type of leadership, sufficient resources, etc.) may be crucial for the development of IR capacity. I will begin with an elaboration on the first set of factors. Massification, which is a rapid growth of demand for higher education, introduced a number of changes that affected both the environment and organizational structure of universities. Among these changes are the growth in public interest in higher education, an increasing number of students, professors, and departments, and an increase in the goals and tasks expected of universities (Altbach, Reisberg, and Rumbley 2009). Because of the expansion of existing institutions, the process of university administration became substantially more complicated, which led to the necessity of a thorough analysis of key processes and groups. In other words, universities became complex and multifaceted systems that are difficult to manage without knowledge of how they work, change, and react to specific decisions. Furthermore, the problems of acquiring financial resources and attracting students and professors, in addition to research and education management, require not only statistical information but also interpretation and analysis of the data in accordance with the strategic goals of the university. The lag of IR in Western European and Asian universities behind the USA is partially explained by the later expansion of higher education in these regions (Neave 2003). The massification of higher education has led to a revision of the relationships between higher education institutions and the state. The former could not fully rely only on public finance and had to develop market strategies for generating revenue. The latter began to serve as an ‘evaluative state’, imposing strict requirements on universities with regard to their financial and institutional efficiency (‘do more for less’) and the provision of accounts of the efficacy of education and research using quality assurance and accreditation instruments. The marketization and privatization of higher education have changed management strategies in universities, which have begun to focus on cost recovery, higher tuition fees, and university–industry links and to place less attention on traditional service functions (Altbach, Reisberg, and Rumbley 2009). Additionally, institutions of higher education are expected to be accountable and meet certain quality standards. IR is a convenient instrument that has helped institutions maintain a balance between widening access to higher education and quality. On one hand, IR is useful in the process of self-study, planning admissions, resources, understanding the market position, and other measures of internal assessment. On the other hand, IR is widely used for official reports on and accounts of the effectiveness of university programs, departments, and initiatives and how the university administration will improve the quality of higher education. Two additional factors, increasing internationalization and the emergence of global rankings, are particularly relevant for the development of IR in research universities worldwide. Research universities are at the frontier of internationalization and are more deeply involved in global communication and networks than other institutions of higher education (Altbach 2011). Internationalization implies a wide variety of initiatives, such as student and faculty exchanges, branch campuses, cross-border collaborative arrangements, English-medium programs and degrees, international accreditation and quality assurance efforts, and other activities (Altbach and Knight 2007; Knight 2012). The need to address international audiences can be an underlying motivation for developing IR, particularly in world-class ‘aspiring’ research universities. First, research universities that attempt to adopt elements of the US research university model (as one of the most successful in terms of research output and rankings) establish data warehouses and introduce IR as a component of their governance systems. Some newly established research universities in East Asian countries and Russia (Altbach and Salmi 2011) are excellent examples of this strategy. Second, prospective international partners and students often request a variety of information regarding the institution before making a decision to establish partnerships, engage in collaborative projects, or apply for a degree, and this information must be collected and arranged properly. Therefore, the purposes of institutional marketing serve as a motivation for research universities to establish IR units. Third, various initiatives in international accreditation, quality assurance, and benchmarking serve as important driving forces for the development of IR practices at research universities. Increasing internationalization stimulated global competition among research universities for talent and resources. Under these circumstances, university global rankings (World University Ranking by the Times Higher Education or Academic Ranking of World Universities) became increasingly important. Rankings could not be easily ignored by universities, despite their limitations; in many cases, rankings became instruments that stimulated a culture of quality (Salmi and Saroyan 2007). Generally, rankings involve the use of objective or subjective data collected from the universities themselves or the public domain (Salmi 2009). Therefore, rankings stimulate universities to collect, analyze, and provide data both for the ranking agencies and for various audiences that determine the universities’ reputations. This factor also justifies the global development of IR offices. The IR agenda for the research university Research universities present special challenges for IR. As relatively large institutions with a strong emphasis on research and innovation, they require a specific IR agenda design. In this section, the functions and goals of IR in the research university, in addition to the most relevant topics and sources of data, will be explored. I will illustrate the conclusions using cases derived from both the existing literature and my work as a director of the IR office at a large research university in Russia. Salmi’s (2009) characteristics of a world-class university will be utilized to formulate a framework and discuss IR efforts in three crucial domains for research universities: attracting and retaining the best students and faculty, acquiring abundant resources, and facilitating a favorable government structure. It must be noted that the examples provided do not embrace the variety of IR projects or reflect different countries’ peculiarities. However, these illustrations emphasize the manner in which IR can contribute to the development of the research university. Talent concentration is one of the key factors of research university success. Therefore, IR practice at research universities should be focused on the provision of data for decision making with regard to the attraction and retention of the best students and scholars. With regard to the latter, studies on faculty recruitment, teaching load, retention, research productivity, and job satisfaction can be informative for the creation of a more productive faculty environment. Students are also an important group with regard to decision making at the research university. It is necessary to examine the admissions process and effectiveness of institutional marketing, student experience, adaptation of freshmen to the university, research engagement, postgraduate student experiences, student satisfaction, and campus climate. There are many ways in which IR can assist research universities in attracting and retaining the best students and faculty. For example, at the HSE, IR efforts are focused more on student satisfaction and success, although there are also some data and analytics provided on faculty. First, the IR office collects and analyzes data on undergraduate and graduate enrolment, which includes demographics, previous education, how prospective students learned about the HSE, how they prepared for entrance exams, why they chose the HSE, what their expectations from the university are, and other data. A special emphasis is placed on those students who were admitted but rejected the HSE’s offer. These studies facilitate understanding of the competitive advantages of other institutions. The results are presented to the Academic Council of the university, and adjustments to marketing and enrolment management are conducted every year. Second, the HSE has a long-standing tradition (since 2001) of collecting student feedback on their education and research experiences via annual student surveys. There are a number of cases in which survey data facilitated the improvement of student services, quality of teaching, and student research engagement. For instance, special student research engagement programs were developed, a first-year student orientation program was introduced, and libraries and dining halls improved based on data provided by the IR office. There are other global examples from research universities of how IR can be informative for decision making in attracting and retaining good students. International students are important in the research university context. One of the findings of the study performed at the Australian University of Technology by Chandra and Sharma (2008) indicates that higher entry scores made no difference to the relative performance of international students but was important for students from Australia. Therefore, the university must consider different admissions criteria for such students. Studies of international student experiences in US and UK universities emphasize the importance of ‘cultural sensitivity’ in the orientation and support of these students (Delaney 2002; Middlehurst and Woodfield 2007). The other significant aspect is student satisfaction and retention in the highly competitive research university environment. There is evidence from Utrecht University in the Netherlands (Möller 2006) and Monash University in Australia (Shah and Nair 2009) that student surveys play an important role in understanding how to increase student satisfaction in research universities. With regard to retention, a recent study based on an IR project at Lund University (Sweden) shows how the introduction of supplemental instruction for first-year engineering students halved the student drop-out rate (Malm, Bryngfors, and Mörner 2012). The faculty component in decision making can also be supported by institutional researchers. It is well documented for the US context that universities can benefit from studies of faculty recruitment, retention, promotion, and retirement (Teodorescu 2012). The HSE’s IR office informs the attraction and retention of the best scholars by collecting and analyzing data on their teaching and research performance, the problems that faculty members encounter while working at the university, and their expectations and suggestions. Similar to the student survey, an annual faculty survey is conducted (since 2003) in addition to special research projects on teaching load, research performance, faculty services, and other areas affecting faculty members. The project on international faculty job satisfaction is also notable. The HSE is one of the few Russian universities that recruits faculty from the international academic labor market. As of 2012, the university had hired 35 scholars with a PhD from leading US and European research universities. The project, initiated by the first vice-rector, aimed at identifying problems experienced by international faculty members in orientation, teaching, research, and the use of university services. A series of in-depth interviews was conducted to obtain insights into the experience of international faculty. The results were confusing: the manner in which international faculty members viewed the university differed greatly from the perspectives of the domestic scholars. The differences in culture and previous university experiences and the language barrier made it difficult for international faculty members to integrate into the wider university community, which led to an international scholars’ ‘ghetto’ inside the university. To overcome this problem, policy changes and integration efforts have been undertaken. The next important factor for world-class ‘aspiring’ research universities is the acquisition of abundant resources for their development. Because research universities generate revenue from the government, students, alumni, and businesses, it is important to maintain a positive image in the eye of various stakeholders and respond effectively to external requests. This goal is associated with a number of activities. The first set of activities is the provision of statistical information regarding the university to government agencies for the sake of accreditation and accountability. Typically, institutions report data on faculty and staff, research productivity, finances, enrolment, student retention, graduation rates, student aid, and other aspects of institutional performance. Research universities, and particularly public research universities, are typically required to report more data than other types of institutions. For example, in countries such as China and Russia, a group of public research universities was identified by the state, and those universities are additionally supported. In exchange, those universities must provide a significant amount of information regarding their performance, emphasizing research and innovation. The second set of activities is the collection, arrangement, and dissemination of information to various stakeholders of the university, including prospective students and faculty, alumni, the local community, and businesses. The research university environment affects these activities. Institutional researchers in research universities concentrate more on communicating the research capacity of an institution to attract more funding and establish partnerships with businesses. Therefore, specific methodologies and indicators must be developed within IR units. An additional challenge of the research university environment is internationalization. Research universities attempt to reach an international audience and must account for the diversity of ways in which people with different cultural backgrounds interpret information. In some cultural contexts, statistics are more meaningful, whereas in others, statements from authorities are considered more trustworthy (Ferren and Merrill 2012). The process of presenting data is becoming more complex in research universities, and institutional researchers must understand cultural and language differences. The third set of activities that are highly important for world-class ‘aspiring’ research universities is to provide data for national or global rankings. Although most ranking agencies obtain information from the public domain or via special surveys, they often expect universities to cooperate on data collection. Regarding rankings, the reputation and funding of the university is often at stake; therefore, it is critical to provide the ‘right’ data. Cheslock and Kroc (2012) discuss the ethical problem that institutional researchers encounter when data definitions provided by the government or ranking agencies are not clear. The authors argue that ‘it is best to provide multiple views of the data, along with descriptions of the pros and cons of each view to assist with the final decision’ (233). An actual knowledge of ranking methodologies and limitations is necessary because they are not only important in terms of the university’s reputation but also affect the distribution of financial resources among the departments. Finally, the third element of Salmi’s (2009) framework for becoming a successful research university is the facilitation of a favorable governance environment that promotes academic freedom and a supportive regulatory framework. There are several methods for how IR can assist in developing favorable governance. First, IR can stimulate collaboration and horizontal connections and facilitate a broad organizational view among university constituents (Leimer 2009). This role can be useful to research universities because they are complex organizations with a high degree of decentralization and specialization. Leimer’s (2009) study demonstrates that a collaborative culture is an intrinsic feature of IR and promotes interactions among people and departments in large universities. Second, IR can enhance organizational learning capacity, which is extremely important for research universities operating in highly competitive environments (Borden and Kezar 2012). IR projects provide feedback on institutional effectiveness and facilitate changes in programs and policies. It must be noted that world-class ‘aspiring’ research universities are institutions in which change occurs remarkably fast, and IR is an instrument for planning the change and understanding its consequences. Third, IR can be extremely helpful with regard to the implementation of new programs and policies. For example, several years ago, a program was launched at the HSE that was intended to promote academic development and support of young scholars (Chirikov 2010; Pravdina, Skvortsova, and Yudkevich 2010). The instruments of the program include special grants and scholarships, the reduction of teaching loads, and special workshops and seminars. After several years in existence, the department that is responsible for the realization of the program asked the IR office to provide an assessment of the instruments and collect the participants’ feedback. Data collection included the completion of more than 30 in-depth interviews with young faculty members in the program and an extensive review of the best practices utilized by academic and professional development in leading research universities. The project facilitated the improvement of the program and modification of the support instruments. The strategy for working with the program’s participants underwent some changes, and the information strategies were corrected in accordance with the participants’ requests. IR office organization and staffing It was previously discussed that IR can contribute significantly to the development of a research university. In addition, IR office productivity depends on the office’s position within the university structure; the staffing, skills, and responsibilities of institutional researchers; and their level of participation in the decision-making process. I will explore some of these matters and identify factors that facilitate the full realization of the IR potential at research universities. The work of institutional researchers requires a wide range of skills. Terenzini (1993) identifies three interdependent tiers of organizational intelligence that are necessary if an individual is to become a good institutional researcher. The first tier is technical/analytical intelligence, which can be acquired primarily through formal education. This intelligence includes familiarity with the basic definitions and categories of higher education, knowledge of quantitative and qualitative research methods (research design, sampling, statistics, etc.), and computer skills. The second tier includes issues intelligence and addresses knowledge of the university’s decision areas, including enrolment, budgeting, assessment, physical facilities management, and decision-making processes (including informal power and reward structures). Such knowledge and skills cannot be obtained only from coursework; some job experience is also necessary. The third tier is contextual intelligence, which involves understanding the higher education environment and peculiarities of the university in which the researcher works. This intelligence requires a deep understanding of the goals, values, and culture of the institution and knowledge of the external factors that affect the university’s performance. This type of intelligence can only be acquired on the job and typically requires several years of work at that institution. Leimer and Terkla (2009) emphasize the importance of social intelligence for IR. The authors argue that this form of intelligence is critical for the IR professional, who usually possesses little direct authority and must rely on negotiation and mediation skills. Despite this accurate portrayal of the qualifications required for institutional researchers, some studies demonstrate that people working at IR offices have a wide variety of backgrounds (Volkwein 2008). Researchers can possess degrees in education, the social sciences, math, science, business, or humanities. On one hand, this diversity inhibits the professionalization of IR; on the other hand, diversity promotes a multifaceted view of the institution, particularly when people with different backgrounds work together. The literature on IR offers a set of recommendations to increase the effectiveness of IR offices and develop issues and contextual intelligence, in Terenzini’s terms. The first recommendation is to stimulate a continuous interaction between institutional researchers and the university’s administration and key departments. For an IR office to properly perform one of its primary functions, namely, providing information necessary for administrative decisions, it must operate in concert with the university’s administration, deans, and other administrative figures. Because of an effectively established feedback system, the IR office staff members are properly informed as to what type of analytics are required to solve the problems faced by the university at any given time. The staff is capable of informing the administration of the results of their efforts and, importantly, can perform continuous monitoring of university operations, which allows them to develop new proposals for research projects and forecasts (Delaney 1997). The second recommendation is that institutional researchers work on collaborative projects, exchange knowledge, and discuss research results. IR offices are one of the elements within the institutional system that address the discussion and resolution of higher education problems at various levels: university, inter-university, local, and national. The task for institutional researchers is to present research results for public discussion and offer them to interested stakeholders. The presentation of research is the foundation for the discovery of practical applications and the initiation of positive change. The third recommendation is to create conditions for the professional development of institutional researchers. The effectiveness of these offices depends on the professional competency of their staff members, as demonstrated by a study focused on the work of IR offices (Delaney 2001). Delaney notes that the higher the professional level of the IR office staff, the more complex the tasks are that the office undertakes and accomplishes; more professionally competent staff members not only prepare analytical reports based on university statistics but also contribute analyses and forecasts that are used for strategic planning at universities (Delaney 1997). The IR office’s position within the university structure differs greatly among countries and institutions. For example, the typical Australian university has an IR unit with staffing from 10 to more than 30 people, which reports to the Vice-Chancellor’s Office or Deputy Vice-Chancellor (Sharma 2010). At Shanghai Jiao Tong University, the Office of Strategic Planning integrates accountability, evaluation, and IR to support the university’s top management team and other departments (Wang, Wang, and Liu 2011). At theHSE (as at some other Russian universities), the function of IR is decentralized and performed by several units: the Office of Admissions, the Office for Undergraduate and Graduate Education, the Office of Strategic Development, and the IR Office. Therefore, it is difficult to formulate a general statement on IR organization. Volkwein (2008) offers an insightful framework addressing the organization of IR offices within various higher education systems that depends on the size and level of development and centralization. Volkwein identifies four types of organizational structures: the craft structure, small adhocracy, professional bureaucracy, and elaborate profusion. The craft structure characterizes a large number of one- or two-person offices whose primary responsibilities are reporting data and dealing with internal data requests. At the developed stage, craft structures are replaced by small adhocracies: offices with two or three employees, a flat hierarchy, and minimal specialization. The position within the university’s organizational structure often determines the type of work that an office performs. The next stage is professional bureaucracy. According to Volkwein, this model is the most effective for IR practice: research is conducted by no more than four people (up to 20 in some extreme cases) with special credentials and experience in the field. These types of offices are characterized by a degree of labor division and a modest hierarchy and can conduct sophisticated research projects. Finally, the elaborate profusion model is characterized by a high level of decentralization, with the function of IR performed by researchers in different units. Although some scholars argue that this model does not provide the most efficient manner of conducting IR (Leimer and Terkla 2009), the elaborate profusion office type is the most common at research universities, particularly at private research universities in the USA (Volkwein 2008, 14). The study of IR offices at universities in the Middle East showed that universities skipped the craft structure and adhocracy models and established a professional bureaucracy model from the beginning (Volkwein, Liu, and Woodell 2012). The question of whether the professional bureaucracy or elaborate profusion model offers a better IR office organization for research universities is intriguing. The proponents of the former model argue that it takes advantage of the economies of scale that are associated with shared expertise, teamwork, cross-training, and methodological diversity (Volkwein 2008). Conversely, a narrow specialization allows university administrations to obtain a more in-depth view of the processes because the staff works closely on a certain problem (enrolment, campus climate, finance, etc.). The determination of how IR will be organized at the university appears to depend on how the university administration views the department. If the department is viewed as an information clearinghouse or source of data, the elaborate profusion model is acceptable. However, if there is a desire to view the IR office as a ‘think tank’ that is aimed towards the collection and analysis of information and formulation of strong recommendations that are then seriously discussed, the professional bureaucracy model provides a better fit. Conclusions/discussion IR is becoming an essential element in the decision-making process in research universities worldwide. The present study shows that IR is a global but highly contextual phenomenon, and its origins vary among countries and regions. The primary factors that contribute to the development of IR are massification, the marketization of higher education, the changing nature of relationships between universities and the government, increasing internationalization, and the emergence of global rankings. I attempted to demonstrate how IR can contribute to the development of a research university and inform decision-making regarding the attraction and retention of the best students and faculty, the acquisition of abundant resources, and the facilitation of a favorable government environment. I also elaborated on the skills required for institutional researchers, which can be grouped into four intelligence categories: technical/analytical, issues, contextual and social. The organization of IR can vary, ranging from a solo researcher attached to the president’s office to centralized offices with layered structures and a division of labor. The choice of model depends on the role that IR is intended to play within the university: an information clearinghouse or think tank. There are additional problems and unanswered questions that must be addressed by further research. First, what is the real effect of IR findings in the decision-making process? There is evidence from both university administrators and institutional researchers that data are ignored in many decision-making situations and other factors, like power relations, academic background or psychological characteristics, are more influential. Bounded rationality (Simon 1991), the incompleteness of human knowledge, is an additional factor affecting decision making. However, these arguments do not disqualify IR (there are many established professions whose necessity was doubted at the beginning), and the main challenge for institutional researchers is to develop new, creative strategies for informing decision making. It is also important to thoroughly examine different types of situations in which data are or are not used and why. Second, what are the most important steps in establishing a successful IR unit at research universities? Institutional contexts are significantly different, and there is therefore no universal recipe for the development of successful IR programs. Common features include an adequate level of competence of the IR staff and their access to the data warehouses. What other factors, such as the location in the university structure, staffing, and internal structure, are important? Third, how can we assess the effectiveness of the IR office in the research university environment? How can we improve the outcomes of IR so that it becomes a reliable navigator on the ‘road to academic excellence’? All these questions are crucially important for further development of IR and research-driven decision making in research universities.

Research has considerable yields for the discipline of nursing, individual investigators, and sponsoring institutions. However, nurses and individual investigators are dependent upon institutions willing and able to sponsor or host the research, because the institution is the recipient or awardee of the grant (Frequently Asked Questions, 2001). Sponsoring institutions are responsible for research and the infrastructure to support it. This article is focused on institutions’ responsibilities to the funding agency, research participants, and individual investigators in realizing the institutional gain from research, and on the infrastructure and resources necessary for successful completion of funded projects. Although the frame of reference is American universities and funding agencies, the principles likely apply in other countries. Beyond meeting their missions for knowledge generation and dissemination, sponsoring institutions benefit from research. In funding year 2000, for example, the five U.S. schools of nursing with the most funding from the National Institutes of Health (NIH) received between $3.7 and $12.4 million in research income from the NIH alone (NIH Support, 2001). Such research grant funds include money for a portion of faculty investigators’ salaries, stipend, and often tuition for students working on research projects, and indirect cost recovery to support institutions’ libraries, laboratories, research administration, and other research services. Very productive faculty investigators may have up to 100% of their salaries paid through research grants. In some research intensive universities, research grants provide more financial support for students than do other sources of financial assistance. Research funding also contributes to a school’s and university’s credibility, visibility, prestige, and rankings, factors important in the recruitment and retention of students and faculty. Institutional Responsibilities To realize research yields, however, institutional grantees must assume the accompanying responsibilities for successfully conducting and disseminating research. These responsibilities include overseeing and monitoring the conduct of the work, access to and availability of the data where needed, responsible fiscal management of the research funds, and providing of appropriate resources and infrastructure (Table 1). University officials assign direct daily responsibility for conduct, oversight, and monitoring of the research to the principal investigators (PIs) and the schools or departments to which the investigators are appointed, while retaining many oversight and monitoring functions as required by federal regulations. University officials and investigators are responsible for completing research as promised in the funded research proposal (Charrow, 1997). If investigators leave an institution before the research has been completed, the institution’s administrative officers, in consultation with the primary funding agencies and PIs, decide whether the project will remain at the grantee institution or be transferred to the investigator’s new institution. Of paramount importance is maintaining the scientific integrity of the research. The decision to retain or transfer a research grant depends on the amount and type of work that remains to be done, the most efficient and effective management of that remaining work, the resources and environment of the investigator’s new university, the availability of a qualified investigator at the original institution who could assume the PI role, and the desires or interests of any local or regional co funders of the project. The university’s responsibility for oversight includes ensuring quality of the data and safety of study personnel and study participants. To ensure the quality of the data, universities often establish policies on the ethical conduct of research, falsification of data, and responsible publication and authorship, as well as procedures and protections for those who report questionable practices or infractions (whistle-blowers). The university’s accountability for protection of human subjects is borne by institutional review boards (IRBs) that establish procedures for obtaining approvals and renewals and ensuring compliance with the three ethical principles identified in the Belmont Report: (a) respect for person: right of self-determination through informed consent, (b) beneficence: “do no harm,” monitor for adverse reactions, and balance risk and benefit, and (c) justice: fair distribution of potential benefits and harms among subjects (Olsen & Mahrenholz, 2000). A university’s responsibility for the ethical care of animals includes setting policies and procedures for review of study protocols before the research begins and the daily care and euthanasia of animals used in research. Ensuring the safety of study personnel (employees) requires university officials to establish policies and procedures to meet (a) Occupational Safety and Health Act (OSHA) standards for workplace safety, including appropriate ventilation, fire prevention, and handling of potentially hazardous materials; (b) Centers for Disease Control and Prevention recommendations about transmission of blood-borne pathogens and other infective agents; and (c) other generally accepted guidelines for laboratory and animal facilities. In addition, when data are collected in the community, safety considerations must include transportation of personnel to data collection sites and methods to judge the safety of the sites and surroundings—the neighborhoods, participants’ houses, or community facilities. The university, as grantee, is accountable for providing access to research data both during and after completion of the work if necessary. University officials may be asked to produce raw data if study results are questioned, if adverse reactions occur, or if the potential for litigation exists. With the Freedom of Information Act, members of the public may request access to data with the following stipulations: (a) the data of interest were produced after the amendments to 45 CFR 74 required by the Office of Management and Budget (OMB) Revised Circular A110 were enacted, and (b) the data were “publicly and officially cited by the Federal Government in support of an agency action that has the force and effect of law” (Freedom of Information Act, 1999). The full implications of this new act are still being determined, but it appears to allow anyone with a genuine interest to have some level of access to the data and unpublished findings of studies funded by U.S. government agencies, including the NIH, National Science Foundation (NSF), and others. The university, as grantee, is accountable for the appropriate and responsible management of research funds (Charrow, 1997). University officials are responsible for monitoring the appropriateness of expenses charged to grants and ensuring compliance with the funding agencies’ rules for obtaining the awarded grant money and moving money from one expense category to another (rebudgeting). To meet its fiscal management responsibilities, university officials generally provide investigators with monthly statements detailing expenses in relation to the grant’s overall budget and discrete categories. University officials provide the funding agencies with progress reports for grant renewals and financial reports detailing expenses within budget categories and the overall budget. Money left unspent at the end of the project period (and any time extensions) must be returned to the funding agency. Because the university recovers indirect costs as the awarded money is actually spent, the institution benefits most when all the money is spent before the end of the project period. Although not explicitly stated in the notice of grant award, funding agencies expect the PI to publish the results of the research (Brooten et al., 1999). Funding agencies often stipulate that any publication or presentation of the work conducted with their funds carry an acknowledgement of the grant and the funding agency. Indeed, future awards may depend on the investigator’s success in disseminating the results from a previous research grant. Universities generally have policies for responsible publication and presentation of research findings, support a public relations department that can help with getting these results to the public, and encourage faculty investigators to publish through criteria for promotion and tenure. The university’s responsibility for dissemination, then, comes from both its mission for knowledge generation and dissemination and the expectations of the funding agency. Institutional Infrastructure and Resources Infrastructure To fulfill responsibilities for successfully conducting research, the university and the school or college must provide the necessary research infrastructure support. The university has a central research office staffed by people with the necessary expertise in scientific oversight, human and animal subject protection, and financial management. In most research-intensive universities, each school or college also maintains its own research office that works in conjunction with the university’s central office to establish and implement appropriate policies and procedures and to monitor investigator compliance. In addition to responsibilities and services before the award, the university’s central research office is generally responsible for (a) monitoring funding agency regulations and federal policies related to research; (b) identifying the implications of any changes for the institution; (c) developing policies or procedures to comply with these changes, and notifying investigators of the new policies or procedures; (d) ensuring compliance with human and animal subject regulations; (e) monitoring and handling conflicts of interest that arise from faculty members’ having substantial financial interests in companies that could be affected by the research findings (Schimmel, 1996; Terry, 1996); (f) negotiating subcontracts with other institutions; (g) meeting the funding agency’s billing, renewal, and reporting requirements; and (h) handling modifications to the grant with the funding agency. The school or college’s research office augments the services and functions of the university’s central research office and handles functions individualized for the specific faculty investigator and the demands and culture of the school. For example, while the university office administrators monitor changes in policies from most frequent and largest funding agencies (usually federal agencies and major foundations), a school’s research office would monitor changes in policies for its specific group of funders, such as specialty nursing organizations and nursing foundations. Although staff in the central office can perform these functions, the size of the university and differences among schools and faculties may limit the ability to meet the mentoring and day-to-day support that faculty from different disciplines need for research success. Close collaboration and coordination of services between the university’s central office and the school’s research office are critical to the success of the research mission. Research infrastructure also includes support service departments that serve the broader university community. The departments of public relations, human resources, purchasing, information services (computing), housekeeping, safety, and maintenance are critical to a research project. Without these departments, the principal investigator could not get publicity for the grant funding and research findings, hire research personnel, purchase equipment or instruments, keep the research project’s computers running, communicate with other investigators, work with potentially hazardous materials, and maintain a safe and clean working environment for study personnel. In addition, personnel responsible for technology transfer activities provide advice and expertise on copyright, patents, intellectual property issues, and marketing innovations. These personnel help link faculty investigators with venture capitalists (individuals and corporations that invest in business start-ups, providing the financial support to bring the innovation to market). They also communicate the university’s policy regarding profit-sharing on income from patents developed by faculty researchers, for example, the percentage of profits that go to the university, the faculty investigator, and the venture capitalists. The technology transfer office may also be involved with pre award contract negotiations. Expertise in this office is especially helpful in contract negotiations in which the funder, often a for-profit corporation, wants to stipulate conditions that are not in the individual’s or institution’s best interests or that have serious implications for the institution’s future research projects. If the funder stipulates that its officers must review and approve all manuscripts or presentations before submission, for example, the funder could prohibit dissemination of findings with negative implications for the funder’s business, diminishing the investigator’s ability to build on his or her own findings (Terry, 1996). Some stipulations about extent of the funder’s ownership of materials, innovations, or knowledge developed during the funded project can give the funder extensive rights that financially or intellectually restrict future use of that material or innovation by the faculty investigator (Sederoff & Sederoff, 1996) and introduce troublesome issues of ownership of future research by faculty colleagues and students. The technology transfer office often has a contract lawyer to represent both the individual’s and institution’s interests, especially regarding rights to current or future use of knowledge developed under the contract. Resources In addition to the support provided by the university and school infrastructure services, investigators require resources specific to the research project and their workloads in order to successfully conduct and complete the research. Important core resources for all investigators are space, time, and mentorship (Brooten, Brown, & Miovech, 1998). Research Space. The need for appropriate space to house the project is often the principal investigator’s first concern when a grant receives a competitive score and is subsequently funded. Although some believe that grants with bigger budgets deserve more space, the amount of space needed to support the work will vary depending on the type of work to be performed, the number of personnel, amount and type of equipment, data-collection procedures (where and when), and type and amount of data to be stored. In articulating space needs, the PIs will need to consider who will occupy the space, at what times of the day, and for what tasks. Projects in which data collection occurs in the hospital or community may need less space in the school than those in which data collection occurs in the school. If participants come to the school for data collection, space is needed that is easy to find and easily accessible physically perhaps with waiting and child care areas. If data are collected by phone, provisions will be necessary for privacy and quiet for interviewers and others who will be in the office at the same time. Research space that is geographically close to the investigator’s office makes supervision of personnel and timely involvement in decision making easier. The data-collection protocol also will influence space needs, including size and amount of equipment, amount and type of data collection supplies (e.g., forms, kits, reagents), and volume of the resulting data. Some investigators need secure storage and office space both at the school and the hospital or community agency where participant recruitment will occur. Other investigators need “wet lab space” with appropriate safety equipment, space with two-way mirrors, or a number of Internet connections. If the data are to be stored outside the research space, security and proximity of the data to the research space are paramount. Storage of research data after completion of the project or the end of the grant funding is important for institutions because of Freedom of Information Act requirements and possibility of litigation. Investigators should also know the school’s policies on initial allocation of research space and retention of space. For example, do those investigators conducting funded research with substantial indirect cost recovery get priority when space is at a premium? How long can investigators retain the space after the study has been completed but the next grant has not yet been approved and funded? Where will the data be stored when the grant monies have ended but publication of secondary study findings is continuing? Knowing and participating in the development of these policies are important in the continued research of faculty investigators. Investigator Time. Conducting research requires the PI’s time, regardless of funding. In some institutions, when a faculty member gets a grant with salary support in the budget, that support generally equates to relief from other responsibilities, including teaching and committee work, commensurate with the percentage of effort funded by the grant. In other institutions where all faculty have part of their effort routinely designated for research, the grant’s salary support may be needed to cover the faculty member’s designated research time before additional relief is provided. This situation increases faculty members’ workload substantially and decreases motivation to conduct funded research (Brooten et al., 1998). When the grant does not provide financial support for the PI’s time, relief time may be granted by the school’s administration. If the institution indicated in the grant application that a percentage of a faculty member’s time would be contributed for grant activities, this pre award commitment must be honored when determining workload. Relief from teaching places demands on the school to cover the courses the faculty member has been teaching. Grants Management. Appropriate grants management is critical to the success of the research. The principal investigator must make decisions about implementing the research project, selecting strategies that maximize the yield from the research, and extending the available funds. Although the PI made many decisions about the scientific conduct of the research in writing the grant application, other decisions about the practicalities of starting the research have to be made. Such decisions include how soon to fill different types of positions, which position to fill first and getting renewed buy-in from the coinvestigators and other project staff. Then the day-to-day operations of the project have to be considered—which telephone number to list as the project number, whether to hire student employees or regular employees, which characteristics for a position are critical and which are nice to have, and whether data collectors report to the project director or the PI. Collaborating or consulting with experienced researchers can help PIs to manage projects more smoothly and effectively (Byrne, Kangas, & Warren, 1996). When a research grant is funded, the PI takes on many of the tasks, demands, and challenges of a mid-level manager. The PI must quickly become familiar with the university’s policies and procedures for getting an account number, hiring study personnel, purchasing equipment and instruments, and paying consultants, and funder’s policies and procedures for grants management. Accurately estimating the amount of time required to complete the steps in the hiring and purchasing processes, to train data collectors in study procedures and the required human subjects content, and to re-establish access to clinical sites is critical to an on-time start of recruiting participants and collecting data. Thus, decisions made in the initial stages of start-up can have long-lasting financial and operational effects. The data collection and analysis phases are probably the most tedious and stressful stages of a research project. Despite the best planning, unexpected issues almost always arise during implementation. Administrators, physicians, or nursing staff may be reluctant to provide access to patients even though they agreed to do so before the grant was submitted. Recruitment of the sample might be slower than anticipated because of a high refusal rate or because a change in health care or reimbursement decreases the number of available participants, necessitating a decision about when additional recruitment sites are needed and which potential sites will yield the biggest gain without increasing the project’s complexity or costing more than the budget allows. Personnel problems—tardiness, absenteeism, lack of follow-through, poor adherence to study protocols—are almost never anticipated, generally occur, and are difficult to resolve. Financial problems such as finding money for an unplanned expense or to open a new recruitment site are also common occurrences throughout the project. Before any changes are made, the PI must consider the aims and intent of the original grant proposal, any special project features or institutional commitments required by the funding agency’s program announcement or request for applications, the need for prior approval before rebudgeting funds or carrying unspent funds forward into the next grant year. As a result of a variety of obstacles a research project’s activities may extend past the grant period. In many cases the funding agency will grant a “no-cost extension.” Although this extension gives investigators more time to complete the work, generally the agency does not award additional funds to cover these activities. In anticipation of needing a no-cost extension, many investigators try to economize throughout the project in a way that does not negatively affect the project, so they have enough money to maintain their personnel to complete the work. Some effective strategies include staggering the hiring of personnel in the beginning of the project, hiring workers on an hourly basis, and negotiating with suppliers for low rates on special equipment (Yeo & Sampselle, 1996). Sometimes students and junior faculty colleagues will participate in a research project for rewards other than money, such as the experience on a funded research project, course credit, or participation in publications and presentations. At times, additional funds from other sources, such as local or national private foundations or intramural “bridge” funds, may be available. After the no-cost extension period ends, important issues are how long investigators may retain their research space without subsequent funding and whether they may maintain a lighter teaching assignment to facilitate completing the analysis and dissemination and preparing the next grant application. Senior research mentors. Experienced researchers working in collaboration with the school’s administrator directly responsible for the school’s research (associate dean for research, director of research, or dean) are the most valuable project and grant management resources for both new and experienced PIs. These people generally have the breadth of experience and perspective to help PIs make the best possible decisions and successfully negotiate confusing and sometimes conflicting funding agency, university, school, and clinical policies, procedures, and politics. By sharing their own experiences, these senior mentors act as advisors, liaisons, and cheerleaders to investigators throughout the research project. All researchers experience problems in their research studies. Senior research mentors can help investigators identify and think through strategies to address situations such as blocked access to study participants, unexpected and emergency repair of equipment, staff performance problems, communication problems between academic disciplines or cultures, and situations of power imbalances. They can provide support and continued encouragement and usually provide insights from their own comparable experiences. Because of their experience and network, they can function as liaisons, helping junior investigators garner additional resources they may need. Senior research mentors can help resolve situations involving power imbalances perhaps at the recruitment site, with other university departments, and sometimes within the school. In some cases, however, additional help may be necessary and should be available from the school’s administrator responsible for research. Having this person accompany the investigator to the negotiation or intervene ahead of time helps to balance the power and allows the investigator to negotiate more successfully and to gain skill and confidence in such negotiations. Conclusions Grant funding brings both benefits and responsibilities to the institution that receives the grant. A variety of infrastructure supports are necessary for a project to be successful, and the types of supports vary with the features of the project. Indicators of successful grants management are: the study’s aims are achieved and the study is completed within budget; results are appropriately disseminated; publicity and media attention are obtained for receiving funding and publishing the findings; the funding agency is informed about implementation problems requiring change and is pleased with the project; and the current project is a springboard for the next project. The institution’s central services provide monitoring and oversight functions to meet its responsibility to the funder and the research participants. The school’s administrator who is responsible for research and senior researchers also have an important role in achieving these desired outcomes for the investigator, the institution, and the funding agency.

There is increasing interest in the integrity of the research process from conceptualization to implementation, dissemination and archiving [1]. Errant behaviors related to the handling of the research process constituting research misconduct are receiving considerable attention in the global scientific community [2, 3]. Misconduct is often defined as fabrication, falsification, and plagiarism (FFPs) [4]. Reports on research misconduct have mainly been from high-income countries such as the United States of America and the United Kingdom [5]. More recently, similar reports have been made from lower and middle income countries such as Nigeria [6], Kenya [7], and Middle Eastern countries such as Egypt, Lebanon, and Bahrain [8]. A meta-analysis of studies in the last decade estimates that 2.9% (95% CI 2.1–3.8%) of researchers report having committed at least 1 research misconduct. In the same meta-analysis, 15.5% (95% CI 12.4–19.2%) of researchers reported having witnessed others commit at least 1 misconduct [9]. The immediate consequence of proven misconduct in published works is a retraction from the journal with its domino effect on all other works that cited the retracted literature [10]. The impact on the global scientific enterprise includes research waste, loss of public trust in the research findings, and misinformed policies that could be harmful to the public [11]. Studies report that research misconduct (RM) is associated with several factors that can be categorized as individual, institutional, national, and global factors [11– 13]. At each level, some factors enable or inhibit such behavior. Individual motivators to engage in research misconduct thrive where the institutional and national structures to prevent, detect, and sanction research misconduct are perceived to be either weak or non-existent [14, 15]. It is recognized that due to an innate tendency to deviant behavior and the need to secure tenure, promotion, or fame and commercialization of research, some researchers will commit research misconduct [16]. To address the challenge posed by this small minority of researchers, institutions, national governments or ministries of higher education need to develop and disseminate research integrity oversight mechanisms and clearly defined sanctions for proven misconduct. National legal frameworks and institutional policies should underpin such research integrity oversight structures and the sanctions proposed therein [14]. Generally, the institutional and national structures that deter engagement in research misconduct are better developed in high-income countries (HIC) while studies show that in low and middle-income countries (LMICs) such structures are either weak or non-existent [17]. In Africa, few countries and institutions have prioritized the development of structures to address the threat research misconduct poses to the scientific enterprise. Among the countries with such systems is South Africa [18]. There is an emerging interest in setting up structures within the East African countries with Uganda and Kenya demonstrating early institutional efforts to address research misconduct [19, 20]. There is also an African regional effort under the aegis of the African Research Integrity Network (ARIN) to create awareness of research integrity. A paper by one of the founders of ARIN outlines the challenges associated with research misconduct in Africa and how institutions and governments in the region could address the problem [21]. There are increasing contributions of authors from Africa on this subject [20, 22, 23]. The emergence of ARIN saw the organization contribute to the hosting in Cape Town, South Africa of the 7th World Conference on Research Integrity in May 2022 [24]. The goal of our study was to explore perceptions of Kenyan ethics committee leaders, top academic and research institutional managers, and leaders at the national research regulatory bodies regarding the occurrence of and institutional capacity to prevent or manage research misconduct. Methods Study design This cross-sectional exploratory qualitative study was part of a three-phase project to develop the capacity of Moi University to prevent and manage research misconduct. This study sought to document perceptions on the occurrence of research misconduct from the perspective of members of ethics committees, the leadership of academic and research institutions, and national regulatory bodies, hereafter referred to collectively as research regulators. Institutions with a tripartite mission including training, research, and extension or simply, institutions of higher learning (universities) were collectively referred to as academic institutions while research institutions were those with the primary missions of research and extension. We purposely selected officials at the highest level in the administration for two reasons. Firstly, to create awareness of RM as a threat to the scientific enterprise in the institutions they lead. Secondly, we believed that participation in the in-depth interviews would stimulate them to reflect on the RM and lead to a buy into the ideas about institutionalizing prevention and management of RM. The study was done in Kenya between June and December 2018. Study sample Our study population consisted of 17 human subject administrators, secretaries, and the chairpersons of Kenyan RECs, five corresponding officials of the National Scientific and Bioethics Committee of the National Commission for Science Technology and Innovation (NACOSTI), Kenya, and the Pharmacy and Poisons Board of Kenya and a purposive sample of five research directors from participating research and academic institutions. Recruitment of study participants At the time of the study, there were 28 research ethics committees and 2 national regulators. Due to the relatively small number of these clusters, we targeted the recruitment of two participants from each of the 30 institutions (28 institutional ethics committees, the National Scientific and Bioethics Committee, and the Pharmacy and Poisons Board). Letters were sent to potential participants to inform them about the study in general and invite them to participate. Through telephone calls, the study coordinator made individual appointments for the face-to-face interviews at the convenience of each specific participant. The appointments were then shared with the qualitative research expert who was in charge of the data collection. Obtaining an appointment with the higher-level respondents was challenging and required multiple requests and reappointments. Due to a lack of interest in participating, non-availability, and scheduling challenges, we succeeded in interviewing a total of 27 respondents. Interview process Research team One colleague, RA, a qualitative research expert with a doctorate in Social Sciences and well-versed in both research ethics and qualitative research methods led the data collection exercise. The qualitative research expert worked with a team of three research assistants all with Master’s level qualifications and previous experience conducting qualitative interviews. Research Assistants underwent a 3-day training on the purpose of the study and study tools. Data collection An interview guide with open-ended and probes was used to collect data. It was developed specifically for the study but with themes derived from the Research Misconduct Questionnaire-Revised (RMQ-R) [25]. The interview tool is attached as Additional file 1, and additionally, the main questions asked during the interview are highlighted in the Results section. All the interviews were conducted in English and audiotaped and field notes were also made. To start the interview, all respondents were provided with a working definition of research misconduct as “deliberate fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results”. Interviews then focused on the participant’s perceptions of the occurrence of research misconduct; the current capacity to prevent, detect and manage alleged research misconduct; and facilitators and barriers to managing research misconduct in Kenya’s institutions conducting research, as a priori themes. The majority of the interviews were carried out at the workplaces of the participants but where privacy could not be assured, the interviews happened in nearby hotels. The median duration of interviews was 34 IQR 25–46 min. The duration of interviews appeared to vary with the participants’ experiences with, and knowledge of research misconduct noted in the institutions they represented. Participants with wider experiences and knowledge shared more during the interviews. To ensure data saturation, we collected information representative of the range of experiences an perspectives relevant to the research question. No repeat interviews were done and neither were transcripts shared with respondents for comment. Data management and analysis The research assistants transcribed the audio recordings verbatim into word documents. The qualitative research expert then reviewed a random sample of the transcriptions for completeness and accuracy. All transcripts were then uploaded to NVivo version 10 for coding. A codebook was deductively developed based on the a priori topic areas emanating from the study tool [26]. The researchers and the research assistants then reviewed the codebook and incorporated their input into the final codebook used to code all the study transcripts. Field notes augmented the data from the transcripts, as relevant. All coded data was then categorized into thematic areas in line with the study objectives: occurrence, prevention, detection, investigation, and management of research misconduct. The demographic characteristics of the respondents were summarized using descriptive statistics. Illustrative quotes were identified and presented with relevant themes. Reflexivity Two issues may have influenced our review and interpretation of the interview transcripts. First, all the authors are employees of one of the academic institutions from where some of the respondents were purposively sampled. The authors are themselves researchers in this academic setting and are therefore quite conversant with the structures and capabilities therein. This lived experience likely influenced our interpretation and the thematic emphasis in the analysis of the transcripts. Secondly, three of the authors (EW, VN, and JK) are also long-serving members of the local institutional research ethics committee and one author is a member of the National Scientific and Ethics Committee (VN). These members have dealt with suspected and confirmed cases of research misconduct. It is this cumulative experience working in a research ethics committee that led to the project that is partly described in this paper and whose goal was to develop the capacity to prevent or manage research misconduct. Again, this exposure may have influenced our perspectives in interpreting and drawing conclusions from the transcripts. Results We interviewed a total of 27 participants with the majority working on research ethics committees. The participant characteristics are summarized in Table 1. Occurrence of research misconduct To assess the occurrence of research misconduct, respondents were asked: “How common is research misconduct in your view?” Participants agreed that research misconduct was rampant and that this occurred mainly among students as a participant from a research ethics committee stated: “And the percentage is widespread..seasoned researchers in the scale of one to ten, I would give you around three. But for students on a scale of one to ten, I would give you eight (REC 11)”. The participants further opined that students committed research misconduct more commonly largely due to a lack of knowledge of research integrity. An institutional leader further explained that the more established researchers are more concerned about their careers while students just want to complete their studies. The participant said: “So to answer your question I would say I believe it is widespread among the students more than the seasoned researchers. Okay, let me say when I look at academic institutions, you know a seasoned researcher, he or she is may….very much concerned about the publication. My name outside there, I am a renowned researcher in pediatrics so I don’t want to engage in such research misconduct. But if I go back now to the student, for example, my interest will be to graduate. My interest will be just to finish and satisfy my parent and the community (IL1)”. In the opinion of this participant, therefore, the motivation to engage in research misconduct was inversely correlated with the length of the career and the related reputational investment. Current institutional capacity to prevent research misconduct The participants mentioned activities to create awareness on research integrity through seminars and presentations; provision of guidelines and strict supervision of students as mentioned by some of the participants. “Actually, last year, we did at least two seminars on ethics, the importance of ethics, the importance to adhere to them (REC 10)”. Training in research integrity and encouraging adherence to the principles was therefore an important Capacity to detect research misconduct Participants indicated that detection of research misconduct was multifaceted. First, the institutions depend on peer review of proposals by scientific committees. This happens mainly in research centers but some institutions of higher learning also have similar structures. After the internal scientific peer review, proposals are sent for research ethics review where there is an additional opportunity for the reviewers to raise red flags if they detect any signs of plagiarism. In institutions of higher learning use of external examiners to examine students’ theses reports also provides another layer of review where instances of plagiarism can be detected. Some institutions depend on free online plagiarism detection software but a few have custom-made proprietary software such as Turn-it-In R or ithenticate R for deliberate plagiarism scan as one participant said: We have a licensed [plagiarism] software but I don’t know which one they use. But we are required to provide the CDs to any student who will present her thesis or proposals. So, they must submit alongside a hard copy, a CD containing a softcopy of actually what this is and then we ran through that software (REC8). These approaches mainly target plagiarism. Capacity to detect other forms of research misconduct that were mentioned by participants include whistleblowing. A participant said: “Other times when people bring in their studies, we will have probably a whistle blower, calling the secretariat to inform the secretariat about certain aspects that are not really, basically about a study that are not right (REC7)” It was however noted that the whistleblowing process is not formalized. Detection of fabrication and falsification mostly falls under the prepublication peer review system that is commonly at journal level. Capacity to investigate alleged research misconduct There were varied practices for investigating alleged research misconduct cited by our participants including site visits by RECs to interrogate or observe research activities in the field to confirm that research is implemented per approved protocol. A participant said: “We have had such…questionable research practices. We have had an experience where the subcommittee of research actually had to do a site visit (REC7)” Such activities were commonly reported by participants from research ethics committees. The view was however supported by an institutional lead. In the cases where there is a substantive allegation, some institutions have a formalized process with a committee chaired by the deputy vice chancellor for academics which interrogates the allegation. A participant mentioned the existence of a committee in their institution but hinted that such committees were ad hoc. The participant said: “I know they constitute a committee which I think is chaired by the deputy vice chancellor in charge of academics but that is all I know but I would imagine that the institutions have strategies for investigations (REC8)” Capacity to manage research misconduct Overall, our respondents mainly from RECs gave an array of actions commonly taken in cases of research misconduct, although this was focused on student-plagiarized work. The actions ranged from cautioning and correcting the student and asking them to redo the work, to stopping research and even shredding data already collected. A participant from a research ethics committee said: “and with major changes probably, … the researcher has to work on it again then go through the process once more (REC10) In some circumstances, it was mentioned that student disqualification may be recommended. Ethics committees submit annual reports to the national regulator (NACOSTI) and this includes reports on research misconduct. One respondent indicated that cases of research misconduct were escalated or reported to the national research regulator as illustrated by this quote. “If there are issues with misconduct, they are escalated to NACOSTI. (REC7)”. Proven misconduct therefore is handled by a variety of actions including correction, suspension of the training program and escalation to the national regulator. The respondents, however, did not mention what actions the national regulator took on cases of misconduct reported to it. Barriers to management of alleged research misconduct Several barriers were cited by our participants. We acknowledge that they are interrelated and overlapping, but, we present them separately only to highlight the unique aspects of each barrier as pointed out in the quotes. Societal‑level barriers A culture of not following laws or guidelines was identified as an important societal barrier to management of alleged research misconduct as illustrate by one participant. We have a culture of not keeping the laws…not following laws, guidelines (REC17) This implied that even where there is guidance, this might not necessarily be helpful. National‑level barriers At the national level, the lack of a national legal framework that defines research misconduct and provides guidelines for academic and research institutions on the prevention or management of misconduct was considered a barrier. Participants from the national regulator agreed that there was no guidance as exemplified by this quote: “.. I tell you something? At the national level. We don’t have research policy (NR 3)” This was consistent with the impression that it was unclear what action the national regulator takes on cases of research misconduct reported to it annually by institutional research and ethics committees. Institutional‑level barriers The majority of barriers were institutional including inadequate financial support for research ethics, inadequate personnel for structures such as ethics committees as well as supervisors and reviewers of student research. Another barrier was the lack of related guidelines and the dissemination to inform both students and supervisors of expectations. To detect plagiarism, for example, there are dedicated software such as Turn–it–in or ithenticate but these were rarely available in academic institutions. The idea of plagiarism scans was reported to be new and only some institutions have the software as was highlighted by one participant thus: “It [plagiarism scans] is coming up [but]it has not picked up. We are now aware of Turn-it-in and among other soft wares that can help detect (REC10)”. The academic institutions also do not have any peer forum for exchanging views on research misconduct and how to manage it. The challenges of patchy capacity and lack of a common database of all past works in the various universities against which plagiarism checks can be carried out were also mentioned. A participant put it thus: ” … because if people are using the same platform, it will be easy for me to know this work had been approved by [XX ethics committee] or had been rejected by [the same committee]. But, unfortunately the universities have not agreed on a common platform (REC3). Coordinated structures such as databases of previous works and dedicated antiplagiarism software would be essential for curbing plagiarism as the commonest misconduct mentioned. Individual‑level barriers At the individual level, participants mentioned individual laxity and failure to optimally use structures available at their disposal. Additionally, a lack of commitment to quality research work especially through diligent guidance and supervision of students was also cited as a barrier as noted by a participant thus: “But then, if you look at it [student research work] there [are] gaping issues and you are like somebody else went through it for you, you just become a rubber stamping IREC (REC13)”. Participants were evidently pointing to the need to encourage and motivate faculty to diligently supervise student work to entrench good science and research integrity. Discussion In this exploratory qualitative study involving research regulators in Kenya, our respondents perceived research misconduct to be common. Our findings are consistent with the African studies that have assessed research misconduct to be common [6, 17]. Research misconduct was viewed with a focus on plagiarism. Falsification and fabrication of research data were hardly ever mentioned as specific misconduct. Nearly all responses referred to addressing plagiarism, potentially indicating that other forms of research misconduct are neglected. Further, there was a perception that misconduct, specifically, plagiarism was a student problem. Participants did not point out that faculty can also commit research misconduct. Faculty are at risk of succumbing to individual factors associated with the commission of research misconduct due to the pressure to publish for promotion [15]. In this respect, the apparent focus on students to the exclusion of later career researchers is concerning and points to an awareness and capacity gap. Regarding falsification and fabrication, identification requires much greater sophistication in terms of critical appraisal of the scientific literature and being alert to subtle discrepancies that raise red flags about the possibility of misconduct. Detection or suspicion of fabrication and falsification also requires a strong culture of responsible conduct of research (RCR) [24, 27] among members of research teams as well as structures for researchers to be able to report any suspicious behavior for requisite intervention [28]. It also requires robust internal and external peer review processes before and after the submission of a paper to the journals [29] to be able to pick out falsification and or fabrication of data. More recently, there is a move towards Open Science whereby study protocols including analysis plans are published ahead of study dissemination and once the data collection and cleaning is completed, data is locked and made available publicly, allowing other scientists to scrutinize the data whenever queries on research outcomes arise. This concept was adopted and incorporated into the Hong Kong Principles on rewarding researchers [30]. Respondents mentioned some internal peer review activities though the role of Open science was not mentioned. Capacity to prevent misconduct is informal and uncoordinated Prevention of misconduct involves awareness creation about misconduct and its consequences, training in the broad concepts of responsible conduct of research (RCR) and capacity to detect misconduct when it occurs [23]. Indeed, it is part of RCR to report any potential misconduct for investigation and determination. Although plagiarism detection software exists, this study identified that few of the institutions, both academic and research, had subscribed to this software and required students or faculty researchers to provide similarity indices for their works before academic assessment or peer review, respectively. Considering the frequent occurrence of plagiarism and the ease with which it can be detected using dedicated software, the perception that many institutions did not have anti-plagiarism software points to a significant missed opportunity to foster research integrity and prevent misconduct. Managing research misconduct is beyond the mandate of RECs and is hampered by lack of guidelines To manage allegations of misconduct fairly, it is essential to develop mechanisms for investigating allegations, determining outcomes of investigations and sanctioning proven perpetrators. Our study identified lack of such structures, with some institutions using human resource policies to inform research misconduct management process. Given the background of perceived fairly common occurrence of research misconduct in low and middle in come economies [6, 7, 17], our findings point to a lack of capacity. Other authors have also underscored the need for a concerted effort to develop institutional and national guidelines to address research misconduct [29]. Such national guidelines are among the structures recommended for managing research misconduct and which were reported to be missing by our respondents [31]. Among the perceived capabilities mentioned by respondents was the role of RECs in management of research misconduct. RECs, also called institutional review boards (IRBs), have the core mandate of promoting and safeguarding the welfare and safety of research participants. Additionally, RECs are mandated to monitor the conduct of approved research to ensure adherence to approved protocols. While many RECs achieve their first mandate quite satisfactorily, the second is more challenging and is generally poorly implemented across many RECs due to a lack of capacity, especially where many research proposals are reviewed and approved [32, 33]. RECs collaborate with the researchers to adhere to the principles of RCR for the safety and well-being of research participants. On the contrary, when a researcher commits research misconduct, a deliberate subversion of the principles of RCR, the processing of allegations of misconduct is a quasi-legal and adversarial proceeding that is outside the mandate and capacity of a REC. In the US, for example, while RECs/ IRBs may have a role in promoting RCR and, in whistleblowing in cases of alleged research misconduct, the task of managing research misconduct is the mandate of a research integrity oversight office that is designed and empowered to carry out adversarial proceedings for any alleged case of research misconduct [34]. Moreover, the consequences of proven misconduct may include career reputational damage and penal sanctions [35], all of which can only be implemented if underpinned by national legal frameworks. A perception that RECs/ IRBs can be depended on for the management of cases of alleged misconduct, therefore, appears misinformed. Strength and limitations To our knowledge, this exploratory qualitative survey is the first of its kind to explore the perceived capacity existing within Kenyan research and higher education institutions to prevent and or manage research misconduct. The study had one important limitation. The respondents appeared to have varying definitions of research misconduct. Such variations in the definition of research misconduct have been reported by other researchers [36]. Some respondents appeared to confuse research misconduct with the much broader concept of academic misconduct. To achieve consensus, the definition of research misconduct was provided to the respondents. It is possible that the variance in definition affected the respondents’ views on the scope of research misconduct leading to the apparent focus on students. We conclude that research misconduct occurs in Kenyan institutions. However, the institutions do not have dedicated capacity to prevent or manage misconduct, a situation that is contributed to by the lack of national guidelines on research misconduct. We recommend the development of Kenya code of conduct or research integrity guidelines that would cover research misconduct. Additionally, efforts should be made to enhance awareness of research misconduct beyond plagiarism and for all researchers. Investment in and mandatory use of antiplagiarism software would be a simple but productive starting point.

Research is the basis of development in any civilised society. The societal progress largely depends upon the human intellectual endeavors. However, the prosperity of a nation depends on how efficiently the researchers are performing to meet the requirements of humankind. Indeed an assessment of research performance is a basic prerequisite for decision making on possible investments, science governance, and academic administration thereof. So the evaluation of research productivity has become an obvious necessity, and is therefore an integral element of R&D institutions worldwide. A large number of evaluative studies on publishing productivity concerning an institution or multiple institutions (read as universities or departments) have been made by many scholars. Quantitative measures have been used quite often to analyse the scholarly performance of an individual scholar, aggregated into groups of varying size or scientific institute and long studied by the scientometricians. Nows-a-days evaluative scientometrics has become operative also by the researchers from other domains. This paper presents an in-depth review of literature on ‘institutional research productivity’ in a globally changing context. This paper aims to depict the state-of-the-art knowledge relating to the topic. Purposively it describes the evaluative studies emphasizing on Indian vista. It also refers to the developmental activities performed by the researchers incorporating newer ideas achieved over time. Sincere effort has been made to map the quantum of knowledge relating to this emerging area of scientometric research. It is however indicative to find the gaps and shortcomings in this specialty of research; hence clearly pronouncing the issues both attended and unattended. Thus it presents a coherent picture of this perplexing problem in the measurement of science. 2. SCOPE AND METHODOLOGY The paper, being the first of its kind, offers a thorough review on the ‘institutional research productivity’ in two different aspects. Primarily it draws attention to the literature concerning fundamental (conceptual) developments for assessment of research productivity. Subsequently, a detailed review of the institutional evaluations has been provided, where single-institutional studies have received prior importance. Although the scope of this review is highly extensive, but concentrated only on the institutional evaluations inclined to applied scientometrics. Hence it does not cover the literature related to performance indicators (indices) used in research evaluations. As an earlier impression, Vinkler (2010)1 systematically presented the literature on scientometric indicators for evaluating research performance. Wildgard et al. (2014)2 also reviewed on bibliometric indicators for measuring the productivity of individual researchers. Similar other efforts were made by Thelwal and Kousha (2015)3 ; Rijcke et al. (2016)4 where they reviewed on a few indicators and alternative indicators used in research evaluations. More recently, Waltman (2016)5 conducted an intrinsic review on citation impact indicators focusing the counting methods, normalisation techniques and source databases. We, therefore, felt it worthy to produce an unfold area, which is under study. The chosen topic is hence relevant and quite interesting. This review should enable the researchers as well as scientometricians to have state-of-the-art knowledge on the area concerned. Its’ purpose is to analyse the available literature toward conceptualizing the possible issues of concentration in the future studies. Thus it makes a significance to grasp an idea on the institutional evaluations, thereby aids to formulate the design of new researches with caution. Pursuant to the identification of literature considered for this review, multiple searches have been conducted to get precise information from authoritative sources; viz. Library and Information Science Abstract (LISA), Indian Library Science Abstracts (ILSA), Library Literature (LibLit), Shodhganga, ProQuest, Scopus, Web of Science, Google Scholar and Indian Citation Index. So it covered almost all possible sources of information relevant to this study. Itemised list of published literature has been scrutinised thoroughly based on our prior understanding. Further information relevant to the topic has been identified systematically via cross-references. A special drive for locating information related to other activities on the issue (say workshop, seminar, conferences, etc.) has been initiatied.The review ultimately considered more than 120 items that are found to be most relevant. To make it more convenient to use and realizing the standard practice, collected references have been organised in the Zotero reference management tool. However to understand the clusters of knowledge; appropriate items are categorised in different sections (single-institutional, multi-institutional) and oriented them chronologically under each section. Finally, it revealed a lot of information to observe them for critical review. 3. CONCEPTUAL DEVELOPMENTS One of the first writers to suggest scientific publications as a measure of research productivity was William Shockley, a Nobel laureate in Physics; who was interested in measuring research performance among individuals within a group by analyzing their publications. Shockley (1957)6 viewed that in scientific enterprises, a few researchers considerably be creative than others in producing scientific output. Later on, Price (1963)7 in his breakthrough work provided the basis of publication productivity that typically differs among the scientists and highly skewed. He found that about 6 per cent of the scientists produce almost 50 per cent of the publications in an academia. Such a difference in productivity among scientists has been explained by Cole and Cole (1973)8 in two different ways. Scientists enormously differ in their cognizance, ability and motivation to do creative works. Otherwise, “the success and recognition of productive scientists used to make them more productive” – they opined. The concept of ‘sacred spark’ and ‘accumulative advantage’ in the assessment of scientific productivity created ample interest among sociologists and scientific enterprises world over (Allison & Stewhart, 1974)9. Gradually it has become a prominent issue in the sociology of science. So a good number of researchers have evaluated scientific publications for assessing research productivity in much the same way as viewed by Andrews (1979)10, Lindsey (1980)11, Rao (1980)12, Fincher (1983)13, Koenig (1983)14, King (1987)15, Garfield & Welljams-Dorof (1992)16, van-Raan (1993,1999)17- 18, Russell & Rousseau (2002)19, Coccia (2004, 2005)20-21, Vinkler (2006)22, Abramo et al. (2008)23, and many others. Most of the studies are conducted for measuring either countrywide or domain-specific else organisational research performance, often combining with citation analyses. Such evaluations have been made increasingly by quantifiable characteristics correlating quality-weighted values of scientific publications. Lawani (1977)24 viewed the citations as a quality measure of publications duly illustrated the scientific papers of Nobel laureates and entomologists. Andrews (1979)10 gave a theoretical foundation on the organisation of research and studied R&D performance of different research groups of volunteer countries in Europe. Wilson (1979)25 summarily identified numerous other factors can be effective for measuring research performance. Lindsey (1980)11 formed the basis of measuring productivity taking into account of sharing credit for multiple authorships in publications and corresponding citations. Rao (1980)12 examined that the negative-binomial (over many other distributions) typically entails a pattern of scientific productivity under the condition “success breeds success” in a variety of social circumstances. Fincher (1983)13 brought some theoretical insights and practical indices for assessing the productivity of higher learning institutions. Koenig (1983)14 correlated bibliometric indicators to generate a composite score in the assessment of research performance. He also compared the results with expert judgment and developed a method of producing score (normalised by institutional budget, i.e. input/output ratio), as an indicator of research productivity. King (1987)15 outlined the portfolio of science indicators and its methodological developments used to evaluate the research activities. Sen and Gan (1990)26 conceptualised on bibliometric methods for applying them in the productivity measure of scientists. Garfield and Welljams-Dorof (1992)16 revealed the ‘citation’ as an influential indicator of measuring science and technology. Sen (1992)27 devised the idea and method of determining ‘normalised impact factor’ and can be more indicative to judge comparative performance. van-Raan (1993, 1999)17-18 introduced some advanced quantitative methods in assessing research performance and to map the developments of scientific endeavors. He stated that “science would not exist, if scientific results were not communicated; so communication is the driving force of science – thus publications are essential and formed the basis of all scientific endeavors”. Geisler (1994)28 suggested an improvised technique of measuring the performance of R&D organisations employing ‘key output indicators’ combining quality and quantity, thus compared two research laboratories. Budd (1995)29 addressed on institutional productivity of research and scholarship at various levels. Russell and Rousseau (2002)19 suspiciously argued on the availability of reliable data that has to be used extensively for evaluating the performance of research institutions. Coccia (2004, 2005)20-21 proposed a few scientometric models in view of determining the R&D performance of public-funded research institutes. Vinkler (2006)22 identified a new composite indicator (using both quantitative and qualitative techniques) for measuring research performance of scientific institutions. Abramo et al. (2008)23 presented a methodology of assessing research productivity of the academic institutes through a number of input/output variables; like number of researchers, funding amount, special grants, areas of scientific specialty, quantity of publications, level (quality) of contributions, resource allocations, etc. Garcia et al. (2012) evaluated on the research performance of academic institutions having multi-dimensional prestige in a scientific specialty (influential field of research) to produce an aggregated summary of prestige score through a number of scientometric indicators. Besides conventional indicators, Kim et al. (2014) explored the possibilities of using research collaboration (in different levels) as a measure of institutional performance. Huang et al.. (2015) viewed on research collaborations through a systematic comparison of co-authorship in the pre-web and post-web era. They found a steady relationship between the interdisciplinarity, multi authorship and citation impact during the Internet age. Pal (2015) highlighted on the lateral relationship among coauthors of collaborative publications to determine the density (intensity) in collaborations. More recently, numerous socio-academic cultural indicators have also been suggested by national statutory bodies for evaluating the performance of scientific institutions (NIRF, 2015). Certainly one has to use appropriate indicator/s and relevant counting technique for utilizing quality weighted values of quantity in assessing institutional research productivity, as suggested by Li et al.. (2017). 4. MULTI-INSTITUTIONAL STUDIES A number of evaluative studies concerning two or more institutions (departments) have been reported in published literature. Those studies were aimed at measuring research contributions across the departments and/or institutions in various dimensions. 4.1 Global Panorama A pioneering work has been done by Glenn and Villemez (1970)36 for evaluating the productivity of the sociologists of a few American universities. Endler (1977)37 also evaluated research productivity and scholarly impact of Canadian psychology departments. In a similar way, Rushton and Meltzer (1979)38 examined on research productivity (impact) of 31 Canadian Universities. Thereafter Schubert and Barun (1981)39 visualised on publishing performance of 85 Hungarian institutes by approaching scientometrics. Yankevich (1982)40 analysed on publication productivity of selected academic institutions in Soviet Republics. McCallum (1984)41 initiated for assessing the productivity of US speech communication departments. Irvine (1989)42 systematically described the progression of evaluating scientific institutions through a bibliometric analysis of technical universities of UK. Rushton (1989)43 revisited the British psychology departments to assess them scientometrically for a decade long period. Zachos (1991)44 compared the research performance of two university departments in Greek employing bibliometric indicators. Royle (1994)45 examined the appropriateness of using the Institute of Scientific Information (ISI) databases for measuring research productivity. He eventually analysed the publication output of three Australian universities using the Science Citation Index (SCI) and Social Science Citation Index (SSCI). Haiqi (1996)46 analysed on research productivity of key medical universities in China using quantitative methods and techniques. Vinkler (1998)47 calculated the performance indices among the institutions of the Hungarian Academy of Sciences based on a few scientometric indicators. National Research Council (1999)48 reported on mathematical sciences research institutes in the United States through a stringent process of accreditation. Schloegl et al.. (2003)49 demonstrated on various problems (especially data sources) commonly occurs in research evaluations, thereby evaluated two university departments of Austria. A different approach has been devised by Bonaccorsi and Daraio (2003)50 for analyzing scientific productivity of French (INSERM) and Italian (CNR) institutes of biomedical research. Jokić et al. (2006)51 assessed on publication output of hard-sciences researchers of six Croatian Universities, as reflected in Web of Science. Albert et al. (2007)52 aimed at evaluating the performance of the Spanish Council for Scientific Research (CSIC) in Biotechnology, focusing on technology transfer. Valles-valenzuela et al. (2009)53 quantified the international exposure of Spanish Universities in the area of legal and forensic medicine by analyzing their research publication data available in Medline database. Lee et al.. (2012)54 studied the impact of collaboration on publication output of publicly funded research institutions in Korea (via network analysis of Scopus data). Ketzler and Zimmermann (2013)55 conducted a citation-based study exploiting Social Science Citation Index and critically measured the influence of research publications of the German economic research institutes for a decade long period. Smyth and Mishra (2014)56 compared the research productivity and citation impact of in-bred and out-bred faculties employed at 21 law schools in Australia. Anyaogu and Iyabo (2014)57 examined some demographic variables as correlates of research productivity taking into account of Law faculties in the Nigerian Universities. They found a positive relationship between ages (experience) with publication output. Pastor et al. (2015)58 pointed out the difficulties of measuring research output of higher education institutions, summarily proposed a simple indicator permitting both qualitative and quantitative aspects of their influence in teaching and research activities. Guskov et al. (2018)59 initiated a project for enhancing the capability of reinforcing publication-output through a study on measuring the research productivity of top twenty-one Russian universities. 4.2 Indian Vista Scientometric evaluations focusing research output of two or more Indian institutions are also evident. First of its type of study was conducted by Sen (1992)27 in the Indian scenario. He analysed on research publications of the CSIR Laboratories by introducing a technique of normalised impact factor. Munshi (1994)60 initiated a study for assessing the publication output of Indian agricultural universities. Nagpaul (1995)61 examined on the research performance of the Indian universities employing interrelated factors of quantity and quality. He measured the quantity by counting articles published in SCI covered journals (only), and assessed the quality via impact factor (normalised) of the source journals. Subsequently, he compared the institutes using related indices; viz. activity index, citability index, relative quality index. Kumar (1999) demonstrated on the idea of determining scientific performance based on the publications of CSIR Laboratories in India. Dhawan and Gupta (2007) took a serious interest in of measuring the performance of physics-research institutes in India, using INSPEC as their primary source of data. Sevukan and Sharma (2008) evaluated the performance of biotechnology researchers of a few Central Universities in India through PubMed, Web of Science, and National Centre for Biotechnology Information (NCBI) data. Prathap and Gupta (2009) devised a new performance index for ranking Indian universities through a complex procedure of quantity and quality measures on research publications. They conducted another study (Prathap & Gupta, 2011) to rank the Indian medical colleges based on research performance using robust indicators of quantity and quality. Kumar (2010) analysed on scientific publications of the oilseed-crops research institutes under the Indian Council of Agricultural Research (ICAR). More recently Husian and Muzamil (2011) performed a scientometric assessment of Central Universities of India by analyzing their publications available in Scopus. Abilash (2012) put an effort to evaluate on the research output of selected higher learning institutions located in Kerala. Kaur and Mahajan (2012, 2015) , compared the research performance among two premier medical institutes, and also ranked the Indian medical institutes based on their publication output. Bala and Kumari (2013) mapped on research publications produced by the National Institute of Technologies (NITs) over a decade. Gupta et al. (2013) compared the research output of Karnataka University with three other universities in the state. Pandita et al. (2014) undertook an analysis of publications produced by four medical research institutions in India using Web of Science. Satpathy and Sa (2015) measured the research productivity of a few universities of Odisha through a bibliometric analysis of their publications reflected in Scopus. Sangeeta (2016) pursued her effort on in measuring the publication productivity of the academic universities in Punjab. Rosalin (2016) gave a clear picture on of the research productivity of the academic universities in Tamil Nadu. Solanki et al. (2016) viewed on research competitiveness amongst the IISERs (Indian Institute of Science, Education and Research). Mukherjee (2017) reviewed on research performance of the CSIR Laboratories in India. He presented more current state of the art of publications using Web of Science database. Batcha (2018) analysed the research publications produced by top six universities of Tamil Nadu. Pradhan and Ramesh (2018) presented a scientometric map of the research publications of six IITs, as indexed in Scopus. 5. SINGLE-INSTITUTIONAL STUDIES Other interesting efforts on measuring research performance targeted to an institute have been made by many scholars in the national and international scenario. 5.1 International Efforts As an earlier impression, Bindon (1981) analysed the scientific output of the Pulp and Paper Research Institute (Canada) using various quantitative techniques. Irvine and Martin (1985) described the reasons of for the growing need for research evaluations. Subsequently, they measured the research performance of the CERN (European Laboratory for Particle Physics at Geneva, Switzerland). Simeon et al. (1986) evaluated publication output of the Institute for Medical Research and Occupational Health in Zagreb (Croatia) to understand relevant policies for academia. Le-Minor and Dostatni (1991) studied on research performance of the French National Institute for Health and Medical Research in order to develop a tool for scientific decision-makers. Beck and Gaspar (1991) assessed the performance of the five natural science departments of Kossuth Lajos University (Hungary) considering the journal impact factor as a quality indicator. Bradley et al.. (1992) analysed the publications and corresponding citations of the Department of Information Studies at the University of Sheffield. Noyons et al.. (1999) demonstrated on how to set a benchmark for productivity assessment of a particular research institute with cognitive orientation and impact. They have used a combined method for evaluating research performance of Inter-University Centre for Micro Electronics (IMEC) in Belgium. In the present century, similar studies have also been made by many others with improved indicators to obtain better insights. Frohlich and Resler (2001) analysed on publications of the Institute for Geophysics at the University of Texas to understand certain discrepancies in publication-counting methods. Lee (2003) examined on research productivity of the Institute of Molecular and Cell Biology (Singapore) correlating basic inputs (recurrent budget and manpower) of research. Alibeygi (2008) explained some determining factors (rank, age and family profile) of measuring research productivity of Razi University (Iran). Stvilia et al. (2011) evaluated on publication productivity in the light of collaborative efforts of scientific teams at the National High Magnetic Field Laboratory (NHMFL), USA. Pudovkin et al. (2012) made an assessment of research productivity at the Deutsche Rheuma-Forschungszentrum (DRFZ), a German medical institution. They also compared the citedness among scientists using citation indexes. Kim (2014) studied on the research performance of the School of Biological Sciences at Seoul National University (South Korea). Haq and Fouzan (2017) evaluated the research outcome of the King Abdullah International Medical Research Centre (KAIMRC) at King Saud bin Abdul Aziz University for Health Sciences (KSAU-HS), Saudi Arabia.They retrieved the dataset from the Web of Science and typically employed some bibliometric indicators. 5.2 Endeavors Made in India A large number of studies on assessing scholarly productivity concerning an institute of India has been reported. An earlier effort was made by Garg and Rao (1988) for evaluating scientific productivity of the National Physical Laboratory (NPL) of India correlating some input output indicators viz. manpower, annual budget and research publications. Seetharam (1997) carried out an exercise on science indicators for analyzing the publications of the Central Food Technological Research Institute (CFTRI), Mysore. Jeevan and Gupta (2001, 2002) , analysed the performance and impact of research papers produced by the Indian Institute of Technology (Kharagpur) and compared R&D performance amongst the departments. Consequently, Mehta (2005) put her efforts on measuring organisational productivity keeping in view of research publications of the National Chemical Laboratory (NCL), Pune. Kademani et al. (2005) also analysed the research performance of chemical scientists at the Bhabha Atomic Research Centre (BARC). Singh et al. (2005) studied on impacts of research contributions made by the Indian Institute of Technology (Roorkee) for a decade long period. Angadi et al. (2006) analysed the productivity and trends in research of social scientists at the Tata Institute of Social Sciences (TISS), Mumbai. Kumbar et al. (2008) assessed on the growth and citation impact of research publications of the University of Mysore. Wadhwa et al. (2008) reinitiated a study on comparing research outputs of the NPL produced in two distinct periods. Bala and Gupta (2009) visualised the influence of research publications of Chandigarh Medical College and Hospital, Punjab. Gradually such evaluations have received momentum with the online access to bibliographic databases. Sarkhel and Raychoudhury (2010) made a quantitative evaluation of agricultural research contributions of the Bidhan Chandra Krishi Viswavidyalaya (BCKV), West Bengal. Kumar and Naqvi (2010) mapped natural sciences publications of the Jamia Millia Islamia University (New Delhi) in different dimensions. Nandi (2010) analytically studied on pure sciences research contributions of the Burdwan University, West Bengal. Mishra (2010) pursued his effort to analyse the research publications of the National Metallurgical Laboratory, Jamshedpur. It is also evident that in recent past single-institutional studies have been made quite often. Most of the studies are focusing mainly on public-funded academic institutions to realise the accountability and to justify possible returns on investment. Jeyshankar et al. (2011) analysed the research publications of Central Electro Chemical Research Institute (CECRI), Tamil Nadu for a period of ten years. Vasistha (2011) investigated on research output of the PEC University of Technology, Chandigarh based on the data available in Scopus. Kaur et al. (2011) evaluated the publications of a government medical college at Chandigarh downloading data from Scopus. Baby and Kumaravel (2012) pursued a bibliometric study on research productivity of the Periyar University, Tamil Nadu using publication data from Scopus. Savanur and Konnur (2012) studied on quantitative growth of publications of the Bangalore University in terms of Web of Science. Kumar and Dora (2012) analysed the research performance of Indian Institute of Management (Ahmedabad) based on the publications indexed in Scopus and Web of Science. Sudhier and Priyalakshmi (2013) evaluated on publication trends of the Central Tuber Crops Research Institute (CTCRI), Kerala. Maharana and Sethi (2013) conducted a bibliometric analysis of the publications of Sambalpur University, Odisha as covered in Web of Science. Maharana (2013) further analysed the publications carried out by researchers of the Orissa University of Agricultural Technology, Bhubaneswar as indexed in Scopus. Rautaray et al. (2013) quantified the research contributions of the KIIT University, Odisha based on the data available in Scopus. Kumbar and Gupta (2013) assessed on publication output and citation impact of the Karnataka University in the field of Science and Technology. Baskaran (2013) studied on the research performance of the Alagappa University, Tamil Nadu. Visakhi and Gupta (2013) analysed the research publications of IISER, Mohali as reflected in Scopus focusing on publication growth and citation impact. Wani et al. (2013) examined the publication productivity of the Indian Institute of Technology, New Delhi using Scopus database. Another attempt was made by Chaurasia and Chavan (2014) to analyse the publishing productivity of the Indian Institute of Technology, New Delhi based on the Web of Science database. Anilkumar (2011, 2014) , conducted studies on productivity of the Physical Research Laboratory, Ahmedabad to understand the research trends duly used to allocate the funds and resources. Gopikuttan and Aswathy (2014) viewed on scientific performance of the Kerala University. Pathak and Bharati (2014) quantified the publications of the Botanical Survey of India over a period of thirty years. Leema Helen (2014) understood publication productivity of the Madurai Kamaraj University through a scientometric study. Ghosh (2014) made his intrinsic efforts on research publications of the Indian Institute of Chemical Biology, Kolkata. He summerly realised the growth and orientation of research focus over the decades of the CSIR-IICB. Gautam and Mishra (2015) evaluated on scholarly performance of the Banaras Hindu University, Uttar Pradesh based on the Indian Citation Index. Jeyshankar (2015) examined the research productivity of Indira Gandhi Centre for Atomic Research (IGCAR), Chennai using publication data from Scopus. Duraipandi (2015) pursued his dissertation work to map the research contributions of the Jawaharlal Nehru University (JNU), New Delhi. Siwach and Kumar (2015) investigated on publishing performance of the Maharshi Dayanand University, Haryana retrieving dataset from Scopus. Tripathi and Kumar (2015) put their intrinsic effort on identifying decadal changes in the research output of the Jawaharlal Nehru University, New Delhi based on the publications as available in Web of Science. Khan and Ahangar (2015) thoroughly studied on research profile of the Government Medical College Jammu through Scopus data using some bibliometric indicators. Over the last few years, the assessment of research productivity has begun to flourish by Indian scholars. Nongrang and Laloo (2016) analysed the research contributions of the North-Eastern Hill University (NEHU), Shillong in the field of Biochemistry combining publications from institutional repository and Web of Science for a period of ten years. Mandhirasalam (2016) exercised both quantitative and qualitative indicators to analyse the research contributions of PSG College of Technology, Coimbatore as reflected in Scopus. Naika (2017) put her efforts on measuring the research performance of the Indian Institute of Technology (Bombay) based on publication output reflected in Scopus. Nongrang (2017) initiated a bibliometric inquiry on published contributions of the NEHU, Shillong for a decade long period. Khanna et al. (2017) analysed on Physics and Astronomy publications of the Guru Nanak Dev University (Amritsar), Punjab as appeared in Scopus database. Built upon previous studies repeated efforts are extended the frontiers of knowledge in applied scientometrics. Kumar (2018a) evaluated on the research performance of the Aryabhatta Research Institute of Observational Sciences (ARIES), Nainital for a period of fifteen-years using Web of Science. Subsequently, he revealed on publication productivity of the Tata Institute of Fundamental Research, Mumbai through Web of Science database (Kumar, 2018b). Mondal and Raychoudhury (2018) carried out a performance evaluation of the Saha Institute of Nuclear Physics, Kolkata based on Web of Science dataset. Nishavathi and Jeyshankar (2018) analysed the publication records of All India Institute of Medical Sciences (AIIMS), New Delhi as appeared in Scopus. Mulimani and Hadagali (2018) scientometrically analysed the contributions of the Indian Institute of Toxicology Research (IITR) for a period of 25 years. Bhakta and Bhui (2018) visualized the research performance of the University of Petroleum and Energy Studies (Dehradun) in terms of the publications reflected in Scopus during a decade. Similar other studies may also be available in due course of time. Keeping in mind the huge amount of literature, the authors have considered a few studies with their hierarchy of relevance. Further conceptualisation on the reviewed literature has been made in a three-dimensional way (data–documents– duration). It has been found that, evaluative studies varied with the data source, documents considered, and durations covered. Most of the evaluations have used the publication data from any of the readily available databases (like Scopus or WoS, seldom done by others viz. INSPEC, MathSciNet, PubMed, and Indian Citation Index) without scrutinizing and validating the dataset. 6. OTHER SCHOLARLY ACTIVITIES Besides the considerable volume of literature, a good deal of scholarly activities on the issue is also evident. As such the conferences and workshops have been initiated regularly. For instance, the ‘International Conference on Scientometrics and Informetrics’ and ‘International Workshop on Webometrics, Informetrics and Scientometrics’ are being held regularly. In this regard, the National and international forums are also prominent. International Society for Scientometrics and Informetrics (ISSI); and Institute of Scientometrics (in India) has come into being. Plenty of journals in LIS field cover scientometric studies in their scope. Newer journals are also coming up (viz. Journal of Scientometric Research). Study circles, round table meetings and online web-forums are actively discussing on evaluative scientometrics to realise the composite indicators and scholarly metrics for research evaluation. In India, an earliest attempt in this regard was undertaken by Documentation Research & Training Centre (DRTC) of the Indian Statistical Institute (ISI) through a Seminar on Bibliometrics, held at Bangalore in 1969. The next one also organised by DRTC at Bangalore in the year 1981, and the third event was again held at Bangalore as Fifteenth Conference of the IASLIC in 1985. However the first workshop on “Scientific Communication and Bibliometry” was held at Calcutta in 1988. A landmark event, namely ‘Third International Conference on Informetrics’ was held at DRTC (Bangalore, 9-12 August 1991) organised by IK Ravichandra Rao under the Chairmanship of Jayanta Kumar Ghosh (then Director, ISI). Another important workshop (regional) on `Informetrics and Scientometrics’ was organised by DRTC in coordination with the National Information System for Science and Technology (NISSAT), New Delhi scheduled at DRTC, Bangalore during 16-19 March 1998. The workshop suitably selected the theme as ‘scientific productivity’ for blending research and practices. In 2004, the Central Library of IIT (Roorkee) organised the “First International Workshop on Webometrics, Informetrics and Scientometrics (WIS)” as a part of the Fifth COLLNET Meeting held during 2-5 March. Nowadays, there are many conferences, workshops and hands-on training programs mes. Such as Third-WIS held at New Delihi (in 2007); Sixth- WIS at Mysore (2010); Eleventh-WIS at New Delhi (2015); and National Conferences of the Institute of Scientometrics are being held annually since 2013. More recently, an effort was mooted by ISI for a “National Workshop on Using Different Metrics for Assessing Research Productivity” held at New Delhi in 2012. UGC sponsored “National Seminar on Advancement of Science through Scientometrics” was held at the Department of LIS, Annamalai University in March 2015. UGC-INFLIBNET Centre regularly organises advanced level trainings on “Bibliometrics and Research Output Analysis”. Almost all the LIS schools are covering research evaluation metrics in their course contents. Nonetheless, scientific evaluation of the institutions has become a national agenda today (NIRF, 2015)34). 7. CRITICAL APPRAISAL Aforesaid discussion entails that many scholars (scientometricians as well) have put forth their striving efforts on assessing research performance. However the evaluative studies on institutional productivity have been made quite often world over. Such (single-institutional) studies are more prevalent in India, immensely organised focusing public funded institutions (include universities) of the country. Most of the studies have considered either readily accessible bibliographic databases or citation indexes (Scopus and/or Web of Science) as their source of primary data for evaluations. In reverse, institutional studies are rarely done by exploiting comprehensive dataset, instead of relying only on readymade databases; which needs intrinsic efforts for gathering and validating publication data combining all possible sources. Moreover, the vast majority of evaluative studies were focused primarily on natural and applied sciences research with a very few on social sciences. In most of the studies, quantitative measures have been taken into account for analyzing the publications; seldom done with quality indicators. While some other studies have been made in a crude and rudimentary manner paying very minimum attention to the requisites of evaluative scientometrics; neither validated the dataset, nor used appropriate tools and techniques – rather overlooked or distorted the scientometric arguments. Sometimes they are far beyond the formalities of scientific writings and reported in questionable journals. Such a practice is more prevalent in India during last few years, perhaps due to superfluity of predatory journals, prompted by policymakers. It has also been observed that in many studies, bibliometric methods are used repetitively without having the proper context. In moral, no single method can be sufficient for all cases and no single indicator can work well (equally) in many situations of evaluating R&D performance. Careful attention is indeed essential to employ the most appropriate indicator using objective metrics. However the quality weighted dimensions of quantity always be effective in evaluating institutional productivity. This review has permitted to map the knowledge of institutional research productivity with broad generality, and in particular to single-institutional studies made worldwide. In this juncture, the role of scientometricians is worth mentioning; when Indian scholars have played a crucial role in extending the frontiers of knowledge in bibliometrics and scientometrics. This work could be useful to track many other relevant issues by conceptualizing the ideas expressed herein. Thus it paves the way for improved organisation of research in this area and could be a driving force in producing better research. 8. CONCLUSIONS The most fundamental idea in the socialisation of science is the publication of research results, which allows researchers to exchange thoughts and reliably receive critical responses on their work. However a researcher acquires recognition, subsequently achieves reputation, thereby fulfills esteem value through publishing. Promotions and positions in the academic world are usually determined by scholarly behavior and research outputs. Quantification of research is therefore an obvious necessity in many academic pursuits. scientometric measurements have been recognised as an indispensable tool for intelligent judgment of research activities and scientific behavior of the institutes. None-the-less, research publications found always the best available basis for evaluating research productivity, often combining with influence factors via citation counts. Certainly, the intellectual development in this specialty of research is extensive, but there is an obvious need for cognizance in exercising appropriate metrics. Further, this review observed no such indicator is sensitive enough for assessment of publications in the interdisciplinary research areas. Although metrics-based evaluations never can replace the peer-review process, but can be a compleiment to human judgments; yet scientometric evaluations will continue for the institutions to formulate strategies and evidence-based management policies.

Interdisciplinary collaboration among researchers generally increases productivity, generates higher impact work (Wutchy, Jones, & Uzzi, 2007), and results in the training of more collaborative researchers (Hampton & Parker, 2011). In light of the mounting evidence of the benefits of collaborative research (e.g., Adler & Stewart, 2010; Beaver, 2004; Jones, Wutchy, & Uzzi, 2008; Lee & Bozeman, 2005), it is not surprising that collaboration is increasing across all research disciplines ( Jones et al., 2008; Wutchy et al., 2007). Funding agencies and programs are following suit: because research clusters and teams generate high-impact knowledge and research that contribute to solving big open questions, the last 5-10 years has seen an increase in big-ticket research opportunities for team-science (Halliwell & Smith, 2011). Canadian examples include: Canada First Research Excellence Fund ($1.25B CAD since 2012; CFREF, 2017), Networks of Centres of Excellence ($560M CAD since 2012; NCE, 2017), Canada Foundation for Innovation (CFI team grants, $1B CAD since 2012; CFI, 2017), and dozens of intermediary team/partnership grants through other federal programs. Similar programs are found globally, e.g., NSF Engineering Research Centers (US), Centres of Research Excellence (Australia), and Horizon 2020 (European Union). In all of these granting programs, foundational components of the evaluation and selection process are the level of excellence of the individuals involved (i.e., traditional research metrics) and the strength and cohesion of the team (e.g., proven track-record of the group’s ability to work together as a team). The Canadian NCE program even requires applicants to explicitly justify the synergies of the team that enable the award to have greater impacts than equivalent grants to individual researchers. The role of institutions in these large-scale programs often seems to be reduced to ensuring compliance, reporting, and providing matched funding for large team grants in the form of cash (e.g., Department, Faculty, and Central funds) and in-kind (administrative and reporting support, space, etc.) contributions. However, for the administration and leadership at an institution to enable faculty to facilitate the creation of truly transformative research programs and therefore to be more successful in these competitions, we need to proactively consider how to develop institutional practices that encourage the development and growth of such research clusters even before particular funding opportunities are known. A recent review of the benefits for, and risks to, individual researchers participating in team grants (Canadian Academy of Health Sciences [CAHS], 2017) called for institutions to increase their support and recognition of team science participants. Indeed, establishing and supporting clusters of research excellence now commonly appears in institutional research strategic plans, in one form or another. However, despite a wealth of literature providing researchers with motivation to participate in team science and examples of previous successes (e.g., Adler & Stewart, 2010; Boardman & Ponomariov, 2014; Guise, Winter, Fiore, Regensteiner, & Nagel, 2017; Reichman, 2004; Stokols, Misra, Moser, Hall, & Taylor, 2008), minimal guidance is available to institutions on developing policies and processes to support the development of interdisciplinary clusters of research excellence. Over the last three years, we have piloted institutional support of the development of research clusters. In this paper, we suggest a framework for identifying, evaluating, and catalyzing clusters of research excellence. We describe and justify our approach, providing specific examples of internal processes and analytical tools that we have implemented and end by discussing challenges and early successes of the program, summarizing lessons learned. We hope that this paper will be useful for other institutions and will spark further dialogue about the roles that institutional administration and leadership can play in supporting research clusters. Planning support for the development of research clusters Collaboration is a central theme in our institutional strategic plan (University of British Columbia, 2018) and enabling the development of collaborative research clusters is an identified core strategy. With this goal in mind, we set out to first understand baseline patterns in collaboration in interdisciplinary areas at the University, and then to identify any existing institutional barriers restricting collaborative approaches to research. In this section, we describe our approaches to those challenges and how these exercises were critical in designing our collaborative research support program. Scoping baseline collaboration in interdisciplinary research areas In larger institutions with thousands of faculty members, a lack of collaborative research initiatives might simply reflect a lack of awareness of other researchers working on related topics in other departments. To assess this issue, we analyzed the extent of pre-existing collaborations among researchers working on related topics and explored whether or not research clusters would develop organically around interdisciplinary topics following strategically designed networking opportunities. We chose four interdisciplinary research areas that had been identified by a recently established cross-faculty consortium whose mandate is to coordinate interdisciplinary health research and education: Indigenous Health, Mental Health, Ageing, and Diabetes: Lifestyle & Biology. First, we needed to identify researchers across the institution that could meaningfully contribute to research in the four interdisciplinary areas. We started by devising a series of descriptive terms that could be searched through our internal researcher indexing system to identify an initial list of researchers working in each area. For instance, in the case of mental health: “mental” OR “psychological” OR “brain” AND “illness” OR “health” OR “wellbeing” OR “wellness”; “psychology”; “psychiatry”. Examples of systems and databases to search when a centralized search function is not available include: institutional researcher webpages, supervisory records for research trainees, research funding and application tracking systems, ethics application databases, etc. Recognizing that even the most thorough search process misses key researchers (e.g., recent hires; researchers who use only technical words to describe their work; clinicians; digital ghosts), we distributed the draft list of potential researchers broadly among unit leaders to help identify any additional researchers. With a revised list of researchers relevant to each cluster, we were then able to gather evidence of existing research excellence and collaborative trends in each of the interdisciplinary areas, instead of by more traditional organizational units (i.e., faculty, department). To assess research activity and impact, we aggregated traditional research metrics on the individuals (e.g., research funding, major awards, publications, citation impact, media attention/reach). To inform which and how many researchers in the interdisciplinary areas had previously collaborated and whether patterns of collaboration were associated with institutional divisions (i.e., collaborations not happening across faculties), we assessed co-publications among the researchers identified. We used co-publications as a convenient proxy for collaboration because pairwise collaboration data can be mined freely from Web of Science or through third party paid software. However, it is Designing support to increase collaboration Our initial scoping of collaborative research activity in the four interdisciplinary research themes revealed that many researchers had already collaborated in these interdisciplinary fields in the absence of formal institutional initiatives, but also highlighted immense opportunity to support additional collaborations within and across faculties. To decide how best to support the development of collaborative research groups, we supplemented the feedback from the four working sessions with broader consultation within our research community (researchers and leadership across faculties) and an environmental scan of support programs at other institutions. Our initial environmental scan of comparator universities found that institutions vary in their definition of interdisciplinary research clusters and consequently in their pathways to identifying and supporting institutionally recognized clusters of research excellence. Most universities define and organize research clusters by disciplines of institutional strength, determined internally or externally (e.g., Simon Fraser University, 2018), while others organize around institutionally identified Grand Challenges (e.g., University of California Los Angeles, 2018) or economic sectors (e.g., University of Toronto, 2018). However, we did not discover a single instance where University support for collaborative research was targeted at providing developmental support for grassroots initiatives and self-organizing research clusters. In spite of this apparent gap, our research community strongly advocated for such an approach and in the absence of model support programs to emulate, we created a novel support program. Through on-going engagement with our internal research community, we settled on a more general (and discipline-agnostic) definition of research clusters as interdisciplinary networks of researchers who organize to solve challenges facing society. Researchers comprising clusters should represent established leaders and rising stars in their areas of expertise working closely together as a unit on complex problems that often transcend traditional departmental, institutional, or disciplinary boundaries. To develop structures and processes that would support the development of such broadly defined research clusters, we described cluster support through a tiered development framework (Figure 1). Figure 1. Conceptual tiered framework for development of research teams from emerging clusters to global leaders, showing (on the left) the characteristics of clusters at various developmental stages and (on the right) the catalytic activities needed to continue on a trajectory to becoming a world-leading research cluster. Our research cluster support initiative was designed to enable these catalytic activities. We first identified characteristics for each tier that should remain true for clusters working on any interdisciplinary research theme and the catalytic activities that would be needed to move from one developmental stage to the next. The catalytic activities identified as essential to cluster development echo the types of support requested by the researchers at the four aforementioned collaborative research workshops. Many of these catalytic activities lie outside traditional academic research funding frameworks (e.g., multi-stakeholder partnership development), are not eligible costs in traditional funding models (e.g., hiring innovation development staff ), and are not widely recognized in reviews of scholarly performance (e.g., community engagement). Without strategic planning and institutional resources supporting these catalytic activities, clusters are likely to maintain current research trajectories and run the risk of not advancing further. Our strategy for supporting the development of clusters of research excellence was therefore centered on supporting these activities. Implementing cluster support programs Securing funding for catalytic activities can limit cluster development when external grant opportunities for smaller, more flexible awards are not easily discoverable, require developed applications, and/or introduce significant time delays before cluster-catalyzing activities can occur. Furthermore, because the funding required for these activities does not include direct costs of research, we anticipated that relatively small awards could have large impacts on the clusters’ development. Balancing the desire to support the development of interdisciplinary research clusters with the recognition that our university cannot support all emerging research teams, we piloted an internal competition to provide small seed grants to self-organizing clusters: Grants for Catalyzing Research Clusters. In this section, we describe our approach to selecting which clusters to support and the rollout of our development support. Identifying and selecting clusters to support While institutions may have well-developed protocols for internal competitions, processes for selecting interdisciplinary research clusters have important nuances that require special attention. For instance, traditional research metrics vary across disciplines, and so aggregate metrics are not often meaningful in the assessment of a single cluster or when comparing multiple clusters. Secondly, it may not be possible to quantify the relative contributions of cluster participants when the group includes a wide range of contributions (theoretical, system specialists, network connectors, etc.) and a variety of roles essential for the cluster’s functioning (e.g., leaders vs. coordinators vs. participants). Furthermore, cluster activities and goals should vary among teams, obviating direct comparison of goals and activities among clusters. Despite these challenges, an evaluation process is required to select which clusters to fund and to evaluate funded clusters over the course of their development. Our approach to evaluation of research clusters has been to focus on broadly-defined criteria where clusters can construct their own cases for fit to criteria, using evidence relevant to their cluster. Example criteria include: • The cluster addresses one or more complex and key questions facing society and has the potential for transformative impact on the University and on society; • Proposal leverages cluster funding to attract further funding opportunities; • Research is interdisciplinary, inter-institutional, and inter-sectoral; • Demonstrated evidence of excellence in research, scholarship, and/or artistic creation; • Demonstrated track record of collaboration and/or teamwork (e.g., co-publications, co supervised students, team grants, etc.); • Evidence of knowledge translation and mobilization activities (e.g., community engagement, policy impact, commercialization); and • Ability to achieve a sustained funding model. Inter-disciplinary panels then review applications and score evidence of fit to the criteria and a strong budget justification that aligns specific activities with goals and expected outcomes. Because applications span multiple disciplines, we ensure that each application receives four independent reviews from researchers in several disciplines and with diverse expertise and perspectives. Reviewer scores are then used to guide an in-person reviewer panel where proposals are discussed among all reviewers and ultimately funding decisions are made. Although the cluster initiative, program call, and selection processes were designed with consultation from our research community, the resulting funding program was dissimilar to models that are familiar to most researchers. Therefore, ensuring success of the initiative would require training of potential applicants (in preparing applications) and reviewers (in selecting applications). To increase the likelihood of generating a competitive pool of applications aligned with the objectives of the program, we hosted University-wide information sessions where we provided details on the program and responded to questions from researchers. To ensure that the review panel understood (and ultimately selected applications that were aligned with) the intent of the program, the review panels started with a presentation and discussion about the intent of the program before discussing applications. Catalyzing development of clusters of research excellence Shortly after announcing the results of our internal competition, institutional research leadership (i.e., Vice-President Research & Innovation, and Associate Vice-President Research & Innovation) met with the leadership team of each cluster individually. During these meetings, we provided high-level feedback from the panel review process in order to refine and focus the clusters’ proposed measurable outcomes and the metrics that the clusters would use to monitor their progress towards those outcomes. These strategy meetings serve as an important link between the review process and the cluster activities, and open up direct communication channels between clusters and institutional leadership to help clusters achieve their goals. Assessing the cluster support program In the pilot year of the *Grants for Catalyzing Research Clusters* competition, we worked closely with clusters to provide guidance when necessary and to learn from the challenges and successes of individual clusters throughout their award period. After the first round of awards (12 months),we also formally assessed the development of individual clusters and the efficacy of the cluster support program by collecting and aggregating post-award reports. Here, we present our evaluation process, early outcomes of the cluster support initiative, and some reflections on the efficacy of the pilot phase of cluster support program. *Evaluating clusters and early outcomes* Because each cluster defined its own goals and expected outcomes in their application to the competition, we based post-award evaluation of their development on their ability to meet self identified goals. Clusters were given a post-award outcomes report comparing their proposed out comes and actual outcomes, justifying deviations. They were also asked to reflect on their experiences and specifically to elaborate on successes enabled by the cluster award and any challenges encountered in developing the research cluster. This information was used internally to evaluate outcomes of the financial investment in the cluster pilot program (i.e., institutional reporting and accountability), to better understand the value of the program from the perspective of the researchers, and to identify opportunities for improvement in the cluster support program(i.e., changes to future competitions). This post-award outcomes report is also attached to future cluster grant applications from the cluster—in addition to their novel proposal being evaluated against the competition criteria, reviewers also rigorously evaluate how well outcomes of previous grants were met. Less than three years from the launch of our cluster support program, we have already observed impact on clusters and on our institution. At the cluster level, we have observed successful ever aging of GCRC funds with federal, industry, and charitable sources, increased collaborations across organizational units (e.g., Figure 2), the formation of new external partnerships, the creation of novel lines of inquiry, and (to our delight) researchers have reported an increased sense of community belonging and interest in collaborative activities. We view these benefits to the clusters as benefits to our institution, but additional institutional-level benefits include:increased external funding, increased partnerships and community engagement, high return on investment for internal resources, expanded networking opportunities for trainees in clusters, increased communication and outreach, and early evidence of significant impact on research. Figure 2. Chord diagram, showing collaborations among faculty members in an interdisciplinary research area (i.e., Language Sciences) across institutional divisions before and after formation of a research cluster. The pre-cluster diagram on the left reveals that most collaborations (as co-authored manuscripts in SciVal) existed within faculties and only a few collaborations existed among faculties. The diagram on the right depicts the pre-existing collaborations plus novels collaborations within and across faculties enabled by the formation of a research cluster (new collaborations are self-reported based on active research projects leading towards publication). The total number of collaborations increased from 47 to 113 and the percentage of collaborations across faculty boundaries increased from 28% to 46%. *Central provisioning of resources vs. providing cash* While the catalytic activities each cluster undergoes ultimately depend on the goals and expected outcomes of the cluster, our initial competition revealed that most clusters share a few fundamental needs, including: communications support, coordination for networking activities, partnership development, funding intelligence, and strategic guidance from institutional leadership. These support needs can typically be best met (in terms of efficiency and quality) through the provision of centrally managed resources. In the second year of our cluster support program, we adopted a mixed support model that provides both institutional in-kind support and cash awards to help each cluster advance. Below, we describe support for developing clusters of research excellence that may be best met through central provisioning. Coordination. Research clusters universally require coordinated activities among cluster members, and with those activities comes increased administrative burden on the researchers. In emerging clusters, this may be limited to organizing quarterly or biannual collaborative working sessions with the larger group and regular meetings with cluster leaders. In these instances, support can be　provided by institutional staff who regularly organize meetings and events or graduate students　involved in the cluster (who may have more bandwidth than faculty members for coordination).In established clusters, the coordination activities needed to keep the cluster running productively may require a full-time staff member dedicated to, or hired by, the cluster. Communications. Developing and showcasing an internal and external narrative is crucial to the success of emerging clusters. Nearly every proposal we received in the first cluster competition requested funding to design and operate a web presence. Yet, it is unrealistic to expect an employee to work with each cluster separately: institutions will not likely have the resources to fund the development of multiple separate websites from scratch, and it is not generally sustainable to bring in a personal communications consultant for each cluster. Additionally, there is no guarantee that the web design and quality will match the institution’s standards. Instead, we provisioned the development of a web template for research clusters from our central IT department and supplied the template to the clusters, saving money and ensuring brand alignment. Our central communications teams then provide communications guidance and support during creation of the clusters’ websites and training of cluster members to support on-going maintenance. Partnership development. In large-scale federal competitions, partnerships across sectors are crucial because they ensure that downstream research users co-create research programs, further leverage funding investment, diversify funding sources, and facilitate knowledge mobilization and commercialization activities. For the same reasons, partnerships are essential to the sustainability of cluster activities. However, even the most highly collaborative researchers may work only in the academic sector. For some researchers that comes as a matter of personal preference, but for many others it may occur because of barriers (actual or perceived) limiting cross-sector exchange. Once clusters have identified their goals and challenges, we are able to connect them with staff experienced in those areas, for instance: partnership development officers (to help with partnership development, innovation plans, and knowledge mobilization pathways [e.g., Phipps, Jensen, Johnny, & Poetz, 2017]); Community Engagement Specialists (in situations where community engagement support is required); Research Funding Development Officers; and Government Relations Officers (when provincial or federal partnerships are key). Established clusters may eventually require their own Strategic Partnerships Officer, but centralized support can get most clusters through the first stages of partnership development (strategic planning and engagement). Cross-cluster exchanges and workshops. In the first phase of the cluster support program, we met with each of the teams to discuss individual goals and strategies to achieve them. The individualized meetings were helpful in the early phase of the cluster support program, but required significant time investment from institutional leadership. General strategies began to emerge; for the most recently funded clusters, we instead hosted a workshop bringing together all of the clusters to collectively (1) set the vision for the cluster support program (2) share successes and lessons learned among clusters and (3) discuss the effectiveness of and suggest improvements to the cluster support program. The workshop still allowed for institutional leadership to help guide cluster development, but also provided the first venue for clusters to interact and learn from each other’s challenges and successes. In a post-workshop evaluation survey, participants unanimously supported the workshop model and have asked that we provide additional programming bringing clusters together to share knowledge in areas of relevance to all clusters (for example: partnership development, governance models, and knowledge mobilization). Top-down vs. bottom-up approaches Our institutional approach to supporting research clusters is to identify self-organizing clusters and support grassroots collaborative and interdisciplinary research teams. Through bottom-up processes, natural leaders emerge and, in our early experiences, this is crucial for sustained cluster management and growth. Bottom-up approaches also allow creative linkages to develop that administration could never have imagined, and would never design, (e.g., a violinist and a climate data scientist) and these are the linkages that generate truly novel lines of inquiry. Bottom-up approaches may also have an added benefit of increased participation by trainees and graduate students (though still a preliminary observation, we speculate this may result from organic relationships among researchers). Finally, supporting self-organizing clusters does not require the extensive proactive background efforts by administration that top-down approaches do (see Scoping baseline collaboration in interdisciplinary research areas section above), and the onus of demonstrating research excellence and the merits of collaborative synergies can rest with the clusters themselves. On the other hand, by identifying and promoting areas of priority, an institution can bring additional opportunities and resources to researchers that might not occur without the institutional branding. For instance, facilitating the development of clusters around external priorities or funding opportunities where they are not organically developing otherwise can bring resources to the researchers and institutions that otherwise would not exist. In this scenario, we see the best role of administration as providing strategic support to help mobilize and support the development of clusters in a given research area. Specifically, administration and leadership can assist with connecting researchers across departments/faculties (as described for the interdisciplinary health clusters above), provide examples of frameworks for collaborative research initiatives, help remove barriers to collaboration identified by researchers, and provide incentives to researchers who wish to develop a research cluster in the area of interest (e.g., GCRC competition). Following our efforts to bring together researchers from the four identified interdisciplinary themes, three of the groups submitted an application to the Grants for Catalyzing Research Clusters competition—two of these applications were successful and have started developing a cluster through our bottom-up support processes. Nonetheless, bottom-up approaches to organizing research clusters present their own challenges. Firstly, researchers are rarely incentivized, financially or through award recognition, to pursue cluster activities (Van Rijnsoever & Hessels, 2011). Consequently, researchers may choose to pursue activities that lead to immediate recognition (Landry & Amara, 1998) instead of activities required to organize and manage research clusters, which may be perceived as detracting from publications, grant writing, student training, etc. Secondly, securing the type of funding that researchers need to support cluster growth is frequently a challenge since most important cluster-organizing activities are generally not eligible costs in traditional research granting programs. Finally, the necessity to secure separate funding opportunities for cluster development may significantly delay cluster development (e.g., application processing times and constrained funding windows). Demes, Murphy, Burt Next steps and future challenges Disciplinary bias The guiding intent of our cluster support initiative is to support interdisciplinary collaboration and catalyze collaborative research across all disciplines. However, our initial competition saw an underrepresentation of applications and funded clusters anchored in social sciences, humanities, and performing and creative arts. To create a more inclusive support program, we worked with researchers and unit leaders from those disciplines to refine the competition call, evaluation criteria, and review process. Indeed, these refinements resulted in an increase in clusters led by researchers in social sciences and humanities and, to a lesser extent, the performing and creative arts. However, we still see an underrepresentation of clusters (and proposals) from humanities disciplines. We are continuing to work with faculty and leadership in the humanities and are piloting pre-cluster support to proposal leads from humanities disciplines, but it is important to recognize that a single cluster support program may not ever be able to be fully satisfy the support needs of all interdisciplinary research initiatives. As we continue to encourage scholars from underrepresented disciplines to participate in and lead research clusters, we have also begun to explore additional funding models that may be better tailored to supporting collaborative work in humanities. We recommend that leadership at other institutions regularly engage with researchers in all disciplines to ensure that collaborative research support programs at their institutions are not inadvertently excluding particular disciplinary expertise. Outcomes reporting In the early phases of designing and implementing the cluster support program, we were not certain to what extent we would be able to measure tangible impacts of the initiative within and over what timescale, and so our outcomes reporting was dependent upon clusters’ self reporting their ability to achieve proposed goals and expected outcomes. On the other hand, continuing to secure internal funding for the program requires empirical evidence of significant return on investment of the seed funding. We are currently designing a more comprehensive reporting process that will include the assessment of whether major goals were met, but will be supplemented by a structured report on discrete outcomes (e.g., leveraged funding, knowledge translation activities, and new partnerships) and a qualitative impact narrative. This expanded reporting, complemented with analyses of institutional trends (e.g., changes in collaborative publications as seen in Figure 2) will facilitate clear articulation of the value of the institutional investment in supporting collaborative research. Moving beyond the pilot phase We now have sufficient evidence demonstrating the success of this pilot program. The next major challenge will be transitioning from a pilot program to an on-going support model at the institution. This will include addressing several unanswered strategic questions, including: How many clusters should the institution be supporting? What is the right balance of support to allocate between newly emerging vs. well-established clusters? Which resources are best provided centrally vs. via funding directly to clusters (e.g., communications support, partnership development, collaboration facilitation)? At what point are clusters expected to no longer rely on central resources for development? Over the next year, we will focus on addressing these strategic decisions in order to develop a plan for the post-pilot phase of the research cluster support strategy. Authors’ Note The development and implementation of the initiatives described above were community efforts and we are indebted to a number of individuals and offices at UBC, particularly: the Associate Deans Research, the offices of the Vice-President Research & Innovation and Provost & Vice- President Academic, and Faculty members leading and participating in research clusters. We also thank Janet Halliwell, Dawn Whitworth, Jonathan Pruitt, and two anonymous reviewers for providing constructive feedback on early drafts of this manuscript, and Dmitriy Ryabika for analytical support in collaboration mapping activities. Funding for our initiatives to support clusters of research excellence has been provided by the UBC Excellence Fund. An earlier form of this paper was submitted to the 2017 Society of Research Administrations International Symposium.

Surgical education has seen a paradigm shift in recent years; the Halstedian apprenticeship model of, ‘‘See one, do one, teach one,’’ has been gradually replaced by evidence based training strategies that focus on skill acquisition outside the operating room.1 Along with these changes, and often driving them, research in surgical education has seen an unprecedented growth.2,3 Nevertheless, while the quality of surgical education research has improved dramatically over this time period,2 most research efforts remain uncoordinated and unfocused. Improved coordination and focus of research efforts could further advance the field of surgical education research. Given that institutional funding for educational research in medicine is scarce4,5 and external grants are few and highly competitive, coordination of research efforts would be of great value to the surgical community because it would provide direct support to the areas of highest need. A current research agenda could guide investigator efforts and allocation of limited funding agency resources to the most pressing areas, thus helping to achieve the goal of better understanding and support of teaching practices.6 Perhaps the biggest limitation of existing education research studies is that they typically originate from single institutions and contain small sample sizes, thus limiting their generalizability.7 Multi-institutional collaborative projects, therefore, have the best potential for advancing the field in a scientifically valid manner. In an attempt to address the challenge of fragmented, single-center studies, the Association for Surgical Education (ASE; www.surgical education.com) created the Multi institutional Educational Research Group to promote multi institutional collaborative projects in surgical education. Here, we report on the first task of this group, which was to focus its efforts by defining the areas of educational research and understanding the challenges associated with the conduct of collaborative studies. The specific objectives of this study were to generate a list of priority topics where multi-institutional collaborative research is needed to advance surgical education and to identify the challenges associated with such research. Methods A modified Delphi methodology similar to that used in prior studies was employed to create the research agenda.8–10 In brief, this methodology uses a systematic process of consulting, collecting, evaluating, and tabulating expert opinion on a specific topic without bringing the experts together. It involves a formal group process in which questions are posed and answered anonymously in rounds. Through exposure to the replies provided, members of the group revise their opinions and eventually converge on consensus. Originally developed by the RAND Corporation to assess long-term trends in science and technology and their anticipated effects on society,11 this method has also been used extensively in the medical field to determine appropriate treatments, facilitate directions in technological innovation, and establish research agendas.6,12–15 Key components to a Delphi process include anonymity, iteration (ie, multiple stages), controlled acquisition of feedback, and analytic aggregation of responses. A particular benefit of this approach is that it can sample the opinion of a group of experts without being overwhelmed by unduly influential persons and can be controlled by appropriate feedback and modification to drive findings toward a group consensus.6,12–15 The memberships of the ASE, Association of Program Directors in Surgery, and Association of Academic Surgeons were asked to formulate and submit up to 5 surgical educational research questions where multi-institutional collaborative projects are most urgently needed (round 1) through an anonymous, Web-based initial survey. Responders were also asked to identify, based on their experience, challenges surgical educational researchers are facing today, strengths and weaknesses of the available literature, and barriers and facilitators of multi-institutional research. The submitted questions were then analyzed, collated, and collapsed by an expert review panel to eliminate redundancy and establish uniform clarity of questions for the second Delphi round. Six members of the ASE Multi-institutional Educational Research Group formed the review panel, which consisted of practicing surgical educators and education researchers. Specifically, the group first reviewed all questions and created question categories. Each member was then assigned 1 to 3 categories to review and combine or reword the submitted questions. Each member was also assigned to revie wand revise, if necessary, the new questions generated by other members. Finally, all new questions were reviewed and finalized by the group before inclusion in the next survey. This iterative process ensured the accuracy and quality of generated questions. In round 2, the collated questions were redistributed only to ASE membership to be ranked according to importance using a priority Likert scale from 1 (lowest) to 5 (highest). Average ratings submitted during round 2 were calculated and the top 40 research questions, along with their mean priority rating, were sent back to the ASE membership for further review (round 3). Responders were asked to rerate the questions using the same 5-point Likert scale; the provided ratings were used to rank the questions in order of priority and create the final research agenda. The relationship of round 2 and round 3 rankings was assessed using Spearman’s correlation to establish rating agreement between rounds. Results Participant demographics for each round of the survey are shown in Table 1. There were no significant differences in participant characteristics between Delphi rounds in regards to personal classification, years in practice, and level of training; in addition, these participant characteristics did not differ significantly from the characteristics of the overall ASE membership. On the other hand, participants of rounds 2 and 3 had more education research publications and a stronger interest in education research. In round 1, 103 responders submitted 328 research questions. These questions covered a broad range of education topics; the topic and question distribution are shown in Table 2. The expert review panel consolidated these questions into 84 unique and answerable research questions. Because of the low overall response rate, we decided to redistribute only to ASE members because of the likelihood they are the group most focused on education research issues. In round 2, 62 respondents rated these 84 questions by importance; mean rating was 3.3 6 1 (range 2.44 to 4.06) on the 5-point Likert scale. The top 40 questions from this round were sent again to the ASE membership in round 3 and 55 members responded. The final ranking order of the top 40 research questions was created based on the ratings received (Table 3); mean rating in this round was 3.4 6 .98 (range 2.83 to 4.23). The top rated research question was ‘‘What are the performance criteria a resident has to meet to be considered competent and before independent practice is allowed?’’ The ratings of rounds 2 and 3 were highly correlated (r 5 .79; P , .01). The top 40 questions were rated overall lower during round 3 compared with round 2 (mean ratings of 3.4 6 .98 vs 3.6 6 1, respectively; P , .001) with 8 questions receiving higher ratings and 32 lower ratings in round 3. Response rates to the 3 surveys varied; in round 1, the response rate was 9% (103 of 1,144 surveyed). Because of this low overall response rate in round 1, the investigator team decided to only redistribute the next 2 surveys to the ASE members because of the likelihood they were the most focused group on education research issues. The round 2 survey was sent to 823 ASE members, 201 of whom opened their e-mail notification for a response rate of 31%. The round 3 survey was sent to 791 ASE members, 185 of whom opened their e-mail notification for response rate of 30%. Seventy-five of 103 round 1 responders (71%) had participated in prior collaborative research projects; of these, 30 had participated in education research projects, 58 in noneducation collaborative projects, and 15 in both. Responders mentioned the institutional culture and practice variability and lack of institutional review board coordination as the most common barriers to multi-institutional collaborative education research, while extensive planning, frequent communication, process standardization, and having dedicated research coordinators were identified as the most important facilitators. Stated weaknesses of the existing surgical education literature included single institutional, small sample studies with limited generalizability, and inadequate scientific rigor, and strengths included the improving methodological quality of studies, the advent of simulation, and the multitude of good ideas. Finally, the main challenges for educational researchers identified were time constraints, limited resources and funding, and lack of recognition for educational research. Comments A modern research agenda for multi-institutional surgical education was developed using a systematic methodology and is presented in this article. The modified Delphi process used to establish this agenda is optimal in that it is structured and transparent, adding validity to the results and the opinions of a large group of participants, which are weighted equally in contrast to a consensus in-person process where typically a few influential participants tend to prevail and can determine the final outcomes.8–10 The top 40 research questions as derived by the ASE membership span a broad range of topics such as performance assessment; simulation; teaching methods; medical student preparation and selection; the impact of work hour restrictions; curriculum development; teamwork and communication; faculty development; and others. It is noteworthy that 9 of the top 10 research questions are related to resident performance assessment and achievement of competency by the time of their graduation. This clearly positions this group of issues at the top of multi institutional education research projects in surgery. This focused agenda in surgical education research highlights important knowledge deficits and may assist researchers in the field in concentrating their efforts by strategically establishing their research programs in the areas of highest need. Furthermore, it may assist reviewers and journal editors in assessing the merit of scientific submissions in light of the perceived need. In addition, this agenda may help funding organizations in allocating limited grant resources to the areas of most need and interest. Importantly, through all these mechanisms, such a research agenda may push the field of surgical education forward by promoting coordinated efforts and limiting redundancy. Finally, it is important to note that as in all science, the content and priorities of this agenda are dynamic this means that as the surgical education evidence base evolves, so will this agenda. Delphi consensus surveys, such as the one reported here, offer a snapshot of current priorities. These should be revisited regularly to provide an overview of the ongoing research progress in the field of surgical education. Limitations of this study include a lowersponse rate to our 3 surveys (9 to 31%). Electronic surveys to a large participant pool, however, are known to have low response rates in the literature. Furthermore, the response rate of rounds 2 and 3 (ASE membership) were significantly higher (30 to 31%) than that of round 1 (9%) proving the point that when the sample size of electronic surveys increases the response rate decreases. Another factor that may have contributed to the higher response rate of rounds 2 and 3 is that ASE members, who are primarily comprised of surgical education researchers, may have had a stronger interest and motivation in study participation. Indeed, respondents of rounds 2 and 3 had a stronger interest in education research and more related publications (Table 1). Another limitation of the Delphi process is the potential for influence of the results by the reviewer panel, particularly when reviewing and consolidating the initially submitted research questions. The collation and inevitable modification of initially submitted questions arguably could dilute or refocus the original intent of the respondents. To minimize this potential problem, the review panel collated questions based on consensus and made every effort to preserve the questions’ original meaning. In addition, the panel was diverse in that it was international and consisted of surgeons and nonsurgeons with a variety of clinical backgrounds and research interests, enhancing the validity of the process. Furthermore, reducing redundancy, combining, and clarifying the questions are requirements of the Delphi process. On occasion, this resulted in combining 2 or more questions into 1 research question, such as ‘‘Which is the best method to identify and remediate residents with poor cognitive, technical, or behavioral skills?’’ In conclusion, an updated agenda for multi-institutional surgical education research was developed using the Delphi methodology. This research agenda may enhance the ability of investigators and funding organizations to focus attention to areas most likely to advance the field and of editors and reviewers to assess the merit and relevance of scientific contributions.

A “gap year” refers to the practice of taking a year out, usually between high school and university or college, to travel, volunteer, work, and generally engage with the world in a way that is exciting and personally transformative. For many young people, an increasingly common part of taking a gap year is volunteering with mission-driven organizations, both to contribute to society and to learn about oneself. In this article, we argue that institutional theory needs a gap year—a period in which core frameworks and insights from an institutional perspective are brought into more intimate contact with individuals and organizations attempting to address complex social problems. We believe this would represent a chance for institutional theory to find itself and learn and contribute in new ways. At first glance, institutional theory appears in great shape and not in need of a reorientating and formative experience. It is recognized as one of the most dominant approaches in understanding organizations (Greenwood, Oliver, Sahlin, & Suddaby, 2008). And yet, much like an overachieving high school student who has spent the past few years in writing essays and studying for exams, institutional research leads a closeted life. It tends to circulate only within arbitrarily defined boundaries, refining its technical prowess, and impressing its appointed judges. This perspective has energized a large and vibrant academic community but is largely unknown to managers and inconsequential with respect to the management of organizations. The insights derived from institutional research remain locked within academic circles—published in journals with limited circulation, written for insiders, with attempts to explore practical implications confined to cursory final paragraphs. The result is a significant disconnect between theory and practice that leaves both the poorer.1 One obvious solution would be for academics to invest in translating their institutional studies by writing for practitioner journals or undertaking consultancy. But we believe this would not be enough. We argue that institutional theory needs to “get dirty”—to move away from largely “hands-off” research approaches and revitalize its connections to individuals and organizations in-situ, both to foster greater relevance, and to reenergize its theoretical development. In exploring the potential to reorient institutional theory, we focus on the emerging conversation around “institutional work” (Lawrence & Suddaby, 2006; Lawrence, Suddaby, & Leca, 2009). This approach to institutional analysis has focused on the practices and strategies through which individuals and organizations intentionally shape the institutional arrangements within which they operate (DiMaggio, 1988; Dorado, 2005; Lawrence, 1999; Maguire, Hardy, & Lawrence, 2004; Suddaby & Greenwood, 2005). It highlights the effort required to create, maintain, and disrupt institutions, as well as the complex relationships between these forms of work (Hirsch & Bermiss, 2009; Jarzabkowski, Matthiesen, & Van de Ven, 2009). Studies have shown that actors rarely engage in only a single form of institutional work and are more likely to be involved in a complex mix of creating, maintaining, and disrupting institutions. Despite its potential, the study of institutional work shares with traditional institutional research a significant disconnect from practice. In this article, we explore how the study of institutional work could both contribute to the construction of new solutions that make a practical difference as well as energize the development of institutional theory as an academic project. Gap years need a destination and a purpose. To provide that for the study of institutional work, we examine its potential integration with “participatory action research”—an approach to research that involves the co-construction of practical knowledge by all participants, through processes of research, action and reflection (Greenwood, Whyte, & Harkavy, 1993; Kemmis & McTaggart, 2005; Wadsworth, 1998). Orientated toward social change, participatory action research encourages its practitioners to think about: How to effectively provoke action by research that engages, that reframes social issues theoretically, that nudges those in power, that feeds organizing campaigns, and that motivates audiences to change both the way they think and how they act in the world. (Cahill & Torre, 2007, p. 205) At first glance, the potential connections between the institutional work approach and participatory action research may seem obscure. As studies of institutional work have academic legitimacy, focusing primarily on the dispassionate search for enduring patterns of social interaction and published in academic journals, participatory action research is openly ideological (Fals-Borda, 2006), motivated to empower participants to transform their lives (Cornwall & Jewkes, 1995), and rarely published in mainstream management journals. And yet these two approaches share some important common ground—both approaches emphasize the need to understand the interaction of social structure and agency in creating conditions for stability and change. Thus, the question that motivates this article focuses on what the results would be if studies of institutional work engaged with participatory action research, either by adopting some of its tenets or by seriously engaging with its methods, insights, and implications. Motivating the Journey: The Questions and Challenges of Studying Institutional Work The challenge we set in this article is to shift the study of institutions and organizations toward relevance and closer connection with the practices of individuals and organizations as they work to cope with and influence their institutional context. We take as our point of departure the study of “institutional work”—“the purposive action of individuals and organizations aimed at creating, maintaining and disrupting institutions” (Lawrence & Suddaby, 2006, p. 215). The concept of institutional work highlights the intentional actions taken in relation to institutions, especially those that are less visible and more mundane—the day-to-day adjustments and compromises of actors as they attempt to create new institutions, maintain existing ones or disrupt institutional arrangements. The study of institutional work highlights the messiness of institutional arrangements (Seo & Creed, 2002), and the importance of agency not only in constructing new institutions (DiMaggio, 1988; Garud, Jain, & Kumaraswamy, 2002; Maguire et al., 2004) but also in maintaining and disrupting institutions (Colyvas & Powell, 2006; Lawrence & Suddaby, 2006; Oliver, 1991, 1992). The study of institutional work, thus, represents a framework with a potential to help connect institutional theory more closely and explicitly to practice, both in terms of understanding the nature of organizational practices and making institutional insights more accessible and available to organizational actors. The study of institutional work has, however, a long way to go if it is to make a significant contribution to practice. Our understanding of institutional work is just emerging—as it stands it is both variable and partial. In relation to creating institutions, although the skilled, effortful practices through which actors attempt to create new institutions have been illustrated in a range of domains, ranging from high-tech innovation (Garud et al., 2002), to complex social change (Maguire et al., 2004), existing research has focused less on the work of creating institutions and more on the accomplishment of institutional creation. So when we have made considerable progress in understanding how institutional entrepreneurs effect institutional change, relatively little attention has been paid to why they do so—what triggers lead actors to engage in work to create new institutions. There are also significant gaps in our knowledge about the work of maintaining institutions. Despite the durability of social and organizational structures as a central, defining feature of the institutional perspective (DiMaggio & Powell, 1983; Meyer & Rowan, 1977), relatively little is known about the forms of work that go into maintaining institutions over time (Lawrence & Suddaby, 2006; Scott, 2001). In particular, why actors disadvantaged by institutional arrangements are seen to not only comply with institutional demands but also actively work to maintain them. There are also questions about the work of actors to disrupt institutions. Although the disruptive potential of individuals and collective actors has long been recognized by institutional researchers (Selznick, 1949), it is often subsumed within accounts of institutional creation. Oliver (1992) identified the erosion or discontinuity of existing institutions as a distinctive process, but this has remained largely unexplored as well as the challenges of embedded action (Battilana & D’Aunno, 2009; Suddaby & Greenwood, 2005): how actors are able to marshal the resources, including the emotional and cognitive as well as material resources, necessary to engage in attempts to disrupt institutions. For the study of institutional work to move institutional research closer to the day-to-day concerns and experiences of organizational actors, it must engage with these sets of questions that are explicitly concerned with the motivations, meanings, and relationships that shape actors’ attempts to engage with institutions. Research focused on institutional work is, however, only emerging and faces considerable challenges. Particularly important to our discussion is that although institutional work demands research that recognizes institutional processes as “local and particularistic, context-sensitive, conflictual and ongoing” (Zilber, 2008, p. 163), existing theoretical and methodological frameworks usually applied to institutional studies may leave institutional researchers ill-equipped to achieve this. Studies of institutions and organizations are typically based on primarily retrospective accounts of “successful” institutional change and are consequently less sensitive to how meanings, practices, and structures are interacted in situ and in vivo (Zilber, 2008). Thus, if the study of institutional work is to develop a more nuanced understanding of the role of agency in institutional dynamics, it requires engaging with research subjects and topics in fundamentally new ways. The Destination: Participatory Action Research Participatory action research begins from a position of close identification with the actors on which it focuses. The individuals who in other traditions might be viewed as “subjects,” a “population,” or a “sample” are understood from this perspective as active, engaged, and equal participants in the research process. Participatory action researchers seek to “get up close and personal” with participants and focus on valuing and understanding their insider knowledge (Park, 2001) that stems from their everyday experience and unique situations (Swantz, 1996). Participatory action research encourages full and active involvement of all participants, usually of a defined community, with the hope that critical reflection will lead to an awareness of their positions in society as well as their own resources (McIntyre, 2008). If successful, participatory action research leads to the development of new practices that deliver substantive change in the conditions of participants (Hall, 2005), through a process sometimes termed “transformative praxis” (Fals-Borda, 1991). Priorities in participatory action research include giving “voice” to the experiences of those who are often overlooked in a society and collaborating with those people to find practical solutions to problems they identify. Participatory action research often occurs as an emergent and local process. Maguire (1993) described, for example, how a participatory action research project grew out of her response to a request on a laundromat bulletin board asking for volunteers to help women and children who had experienced domestic violence. Participatory action research often employs creative methods such as drama (Conrad & Campbell, 2008) and fotonovela (Kirova & Emme, 2008) to raise the consciousness and confidence of marginalized groups. McIntyre (2008), for example, invited a Northern Irish working-class women’s group to tell a visual story of their daily lives using Instamatic cameras. The aim of the project was to raise awareness of and attention to their experiences of Northern Ireland’s political conflict over 30 years and the important roles they had played as “vocal and visible champions of justice, freedom and equality for themselves, their families and for their overall communities” (McIntyre, 2008, p. 3). Frequently carried out in developing countries where the issues of marginalization are more apparent, participatory action research can take many years to build the necessary commitment and trust of local members. Swantz (1996), for example, described spending more than 25 years observing the everyday practices and values in Tanzanian communities to understand how those practices and values might persist as the country modernizes. Although participatory action research has a rich history with significant successes, it is also an approach that faces significant challenges. First, participatory action research has been most successful in dealing with issues that are local to small groups relatively, often temporarily isolated from their contexts, leaving the approach open to the criticism that the emphasis on a “micro-level of intervention can obscure and indeed sustain macro-level inequalities and injustice” (Cooke & Kothari, 2001, p. 14). Consequently, participatory action research has been largely unable to engage with large-scale issues, such as global poverty or climate change (Reason & Bradbury, 2007). Second, the role of organizations in participatory action research is theoretically and practically underdeveloped. In fact, participatory action research is often focused on communities and takes an antagonistic view of organizations that are cast as sources of oppression and exploitation (Cameron, 2007). In the cases in which participatory action research has been adapted for use inside organizations, the result has been a rather technical, action research model (Conrad & Campbell, 2008), which has tended to adopt a somewhat selective view of participation (e.g., Sense, 2006), often within a prescribed model of change set by the leaders of the organization (e.g., Street & Meister, 2004). Looking across these challenges, it seems that participatory action research may develop rich understandings of collective meanings and practices, but it lacks a theoretical lens that would facilitate making sense of them in ways that could lead to more fundamental social change. Common Ground Without a doubt, studies of institutional work and participatory action research make a somewhat unlikely pairing. Although studies of institutional work are primarily focused on engaging an academic community, the overarching goal of participatory action research is to change the conditions of a marginalized group practically (Hall, 1992) or transform an organization (Greenwood et al., 1993). Moreover, legitimacy for participatory action research depends on engagement with participants in ways that allow them to influence the research agenda directly (Kemmis & McTaggart, 2005), whereas legitimacy for the study of institutional work relies on the rigor of its methods and its engagement with theory. Despite these undeniable differences, however, we suggest that the two perspectives share important common elements, which could provide the foundation for extending both lines of inquiry. Three ideas, in particular, connect them: the primacy of heterogeneous agency, the importance of practice, and recognition of situated knowledge (see Table 1 for a summary). Primacy of heterogeneous agency. Both perspectives place a premium on the role of agency in understanding and effecting social change and point to the wide variety of actors and actions this can involve. Studies of institutional work view individual and collective actors as intelligent, creative, and purposive, pointing to their “awareness, skill and reflexivity” (Lawrence & Suddaby, 2006, p. 219). Drawing on DiMaggio’s (1988) discussion of institutional entrepreneurship and Oliver’s (1991) strategic responses to institutional processes, Lawrence and Suddaby (2006) argued that more attention needs to be placed on the intentional activities of actors and their work in not only creating but also maintaining and disrupting institutional arrangements. For its part, participatory action research is committed to replacing the traditional subject–object relationship between the researcher and researched with a subject–subject relationship (Fals-Borda, 1991) in which participants are understood as competent, reflexive, and capable of participating in exploring their social worlds and realizing change (McIntyre, 2008). Both perspectives also share a concern to engage with a broad range of actors who are often less visible in traditional research. The concept of institutional work points not only to the entrepreneurs who create new institutions but also to the “janitors” and “mechanics” who maintain them, the “homeless” who fall outside of normal institutional boundaries and the malcontents who disrupt institutional arrangements (Lawrence, 2008; Lawrence & Suddaby, 2006). Participatory action research is concerned with engaging often overlooked actors but emphasizes the recognition of those without “voice” (Hall, 1993) who occupy peripheral positions in society and are in some way oppressed or disadvantaged. The importance of practice. Both the concept of institutional work and participatory action research highlight the role of practice—routine, legitimate sets of skilled social behaviors (Bourdieu, 1977). This focus provides insights into processes of institutionalization, as well as the need and potential for social change. In both cases, practices are understood as “embodied, materially mediated arrays of human activity centrally organized around shared practical understanding” (Schatzki, 2001, p. 2). For studies of institutional work, this idea shifts attention away from the effects of institutions toward the practices that create, maintain, and disrupt them. The theoretical challenge moves from understanding the diffusion of institutions in understanding the nature of shared practices and their institutional effects (Lawrence & Suddaby, Table 1. Finding Common Ground Institutional Work Participatory Action Research Heterogeneous agency The importance of individual and collective actors and their awareness, skill and reflexivity. Recognition of the ‘janitors’ and ‘mechanics’ in institutional processes that have been overlooked. Participants viewed as competent, reflexive and capable of participating in exploring their social worlds and realizing change. Focused on assisting powerless groups of people, for example, the exploited, poor, oppressed and the marginal. Emphasis on practice In studying the practices that create, sustain and disrupt institutions, wider effects might be explained. Interest in understanding how diverse actors combine on common projects. Understanding practice as the basis for action and change. The development of new shared practices to transform social arrangements. Situated knowledge Actors are embedded in institutional contexts. Agency and practices operate within institutionalized rules. Participants are knowledgeable of their specific situation. Recognizing tacit knowledge can enable actors to engage in substantive change. 2006). For participatory action research, practice represents both the target of change and the basis for action. Rooted in Lewin’s (1946) action research methodology with its emphasis on resolving social problems and conflict (Conrad & Campbell, 2008), participatory action research proponents work to expose the oppressive effects of existing practices and experiment with new, emancipatory practices in an iterative cycle of research, action, and reflection (Kemmis & McTaggart, 2005). Through this process, individuals and groups are encouraged to reflect on taken-for-granted practices critically, resulting in the identification of individual and collective capabilities as well as an awareness of how they are disadvantaged. Recognition of situated knowledge. The importance of agency and practice in both perspectives leads to the recognition of actors as knowledgeable agents whose skills and intelligence are rooted in their unique contexts. The concept of institutional work is explicitly concerned with “intelligent, situated” action (Lawrence & Suddaby, 2006, p. 219), reflecting the ability of actors to tailor their institutional strategies to the specific contexts in which they operate. This approach assumes culturally competent actors able to creatively leverage the sets of institutional rules and resources that structure their day-to-day lives. Participatory action research is similarly sensitive to the local context and the situated intelligence of participants. It focuses on the needs of a particular group or community, with researchers spending substantial lengths of time situated “inside the culture” (Swantz, 1996, p. 124) to understand how participants relate to their contexts, the survival strategies they adopt, and the resources at their disposal. From a participatory action research perspective, participants have unique, often tacit, knowledge of their own situations, and their community’s rules and norms (Coghlan, 2003), which makes them distinctively well placed to organize and implement change. Although participants’ tacit knowledge may be difficult to access and articulate (Collins, 2001), the participatory action research approach emphasizes the value of this knowledge in terms of understanding the creativity and complexity of everyday life and the untapped skills and abilities that may be key to social change (McIntyre, 2008). In a Foreign Land: Studies of Institutional Work Informed by Participatory Action Research A key feature of the cross-cultural and volunteering experiences associated with gap years is their potential to transform the individuals involved—affecting their outlooks, their values, the questions they ask, and the ways in which they interpret their own worlds upon returning home. But what might this look like for the study of institutional work? How might an encounter with the methods, insights, and problems of participatory action research help find answers to some of the questions faced by this emerging perspective? We explore this potential by examining the insights of a set of illustrative participatory action research studies for the three categories of institutional work—creating, maintaining, and disrupting intuitions (Lawrence & Suddaby, 2006). Participatory Action Research and the Work of Creating Institutions Studies of actors attempting to create new institutions have dominated the area of institutional work. Following DiMaggio’s (1988) discussion of institutional entrepreneurship, organizational scholars have invested significantly in understanding how and under what circumstances individuals and organizations are able to effect new institutions (Garud et al., 2002; Greenwood & Suddaby, 2006; Maguire et al., 2004). We have argued that missing in these discussions has been a serious consideration of what triggers actors to engage in the effortful, risky work of creating new institutions. We believe that participatory action research may be particularly well situated to help answer this question and draw on a project that was done in post-war Guatemala City focusing on the creating of a new model of child discipline. The study involved a series of processes, beginning with a survey carried out by 14 local women to explore the factors associated with “‘normative’ disciplinary practices (such as smacking), ‘severe’ [actions] (the use of the belt) and ‘extreme’ violence (burns, abrasions and whips)” (McMillan, 2007, p. 522). The survey evolved into a set of discussions among the local women and academic researchers concerning the survey’s methodology and their findings. McMillan (2007, p. 524) argued that these “discussions marked a critical transition in the project. ‘Maltreatment’ (and extreme deprivation even by local standards) was not ‘out there’ but suddenly visible around them.” As the funding for the initial survey wound down, the 14 local researchers enrolled in a biweekly parenting workshop through which they carried on their individual and collective exploration of the issues around violence toward children. The process then developed in multiple, unexpected ways. Some local women became involved in exploring issues of ethnicity, particularly with respect to the division between mixed-race and indigenous people. The parenting workshops led to the formation of a set of formally constituted women’s organizations, which led to women becoming involved, for the first time, in the civil development of the neighborhood. Finally, nearly every member of the original survey group began literacy and vocational training. Although the story of this project appears primarily about action rather than scholarly reflection or analysis, it provides important insights into why actors engage in the difficult, risky work of attempting to create institutions. Although examining the work of impoverished, politically disadvantaged people may seem idiosyncratic, we believe it has general implications well beyond the context of Guatemala City. Several themes emerge. The first is the importance of protected spaces that allow individuals to engage in sharing and collective reflection on their situation and experiences. McMillan (2007, p. 525) reported that in the first set of parenting workshops, maltreating practices came to be “recognised to be part of the experience of those present,” with participants coming to “recognise that physical violence had eroded their confidence and engendered ‘bitterness’ (amargura) that they sometimes vented on their own children.” These protected spaces highlight the importance of collective action in creating institutions and especially how such collective action can be engendered. Unlike research on institutional entrepreneurship, participatory action research tends not to elevate the role of individuals. Moreover, this study reinforces lessons from social movements with respect to the importance of collective processes that are at least as meditative as they are strategic. It suggests that although resources and skills may be necessary to accomplish institutional creation, engaging in the work to do so may be better explained by intense, emotional collective experiences that forge common understandings and identities. A second theme that emerges is the role of controlled escalation of agency. One example of this is the seeking out of literacy and vocational training by the original local researchers. A second involves the way in which local women came to protect children who they learned were being maltreated: If the situation involved a family member or close acquaintance, they acted alone; if a neigh bour, they generally teamed up with another friend in the group to address the problem. They did not discard the option of informing the courts responsible for child protection, but did so only in two cases of strongly suspected sexual abuse. This was a dangerous step, given the possibility of reprisals by the abuser and an indicator of an increasing awareness and sense of shared responsibility for the protection of children. (McMillan, 2007, p. 527) This example is particularly telling with respect to understanding how actors come to the point of working to create new institutions—new ways of disciplining and protecting children in this case. It suggests that they do so neither alone, nor all at once. It suggests that the institutional work of creating institutions involves sets of practices that are learned, often slowly and out of necessity, and that they are truly “practices,” in the sense that they belong not only to an individual but also to a community of actors. In this case, the community of women itself was transformed, both as individuals and as a collective through their development of shared practices, and institutionalization of new routines and understandings. These processes echo those described by Bradbury and Reason (2003) who argue in their study of participatory action research in The Natural Step that individuals behaved more like “change enzymes” than change agents, both acting as catalysts for change and being transformed in the process. Participatory Action Research and the Work of Maintaining Institutions Although both participatory action research and institutional research have primarily focused on understanding change, the combination may generate particularly nuanced and sophisticated understandings of how and why actors work to maintain institutions, and the consequences of that work. Fine et al. (2004) reported on a participatory action research project in a New York women’s prison that illustrates the potential for such work to shine a light on the motivations, resources, and consequences associated with maintaining institutions. The institution in question was a college program that had been running for 15 years in Bedford Hills Correctional Facility, when a shift in the political climate led to its closure along with more than 340 other prison-based college programs. The reaction among Bedford Hills inmates was “a sea of disappointment, despair, and outrage” that led a coalition of inmates, the prison superintendent, a volunteer, and a local college president to resurrect the program as a private, voluntary consortium dedicated to inmate education. The aim of this project was to assess the effect of the college program “on the women, the prison environment and the world outside the prison” (Fine et al., 2004, p. 102). This study provides an intimate view of how people work to maintain a threatened institution, why they do so, and the potential consequences of this work. A key issue that surfaced in this study is the complex relationship between the work done to maintain the institution in jeopardy (the college program) and the work needed to maintain other institutions. Maximum security prisons are designed to control the daily lives of prisoners. Inmates are not free to leave and are subject to continual surveillance, confined to specific areas, and expected to comply with the instructions of prison staff. Failure to conform can result in further restrictions and punishments. The study suggests that working to maintain an educational facility within this environment requires a variety of work that reinforces this system of control. On an ongoing basis, inmates must conform to prison-control systems to gain and keep access to college programs. For the participatory action research team, working to maintain the college program meant keeping in line with the dominant institutionalized systems of control in the prison, both formal and informal. Inmate researchers needed to exhibit consistent compliance with prison rules to keep their place on the team, whereas also complying with the secrecy and privacy norms of the inmate culture. “An inmate doing research is also a person trying to survive and get out of prison” (Fine et al., 2004, p.112). A significant concern of inmate researchers was the reactions of other inmates: “Our relationships with our peers are a basis for survival. We live in a closed community in which everything is tied together. There is no exit” (Fine et al., 2004, p. 112). Thus, we see that both inmate and Graduate Centre researchers worked to maintain the systems of control within the prison while working to, and sometimes in order to, maintain the college program. The interplay of institutional work focused on the college program and the work of inmate and graduate-center researchers in support of existing systems of control in the prison points to important insights regarding why people might sometimes maintain institutions that appear to disadvantage them. In particular, it suggests that to understand the motivations of institutional work and why actors choose particular strategies, we must account not only for the institutions on which the work focuses but also the system of institutions within which that work occurs. In this case, focusing on either the college program that the inmate researchers were working to preserve or on the formal and informal systems of inmate control they were reinforcing would lead to a partial and misleading understanding of the motivations of the inmates. The institutional work of maintaining the college program and the more mundane work of complying with prison control systems were highly interdependent: Compliance allowed the inmate researchers to engage in the participatory action research project, whereas the participatory action research project, and involvement in the college program more generally, bolstered the motivation of inmate researchers to cope with the indignities of prison life. The project team decided that some aspects of their study could not be shared for fear of possible retribution from prison authorities. If compliance to prison rules resulted in maintaining college programs then that was deemed an acceptable trade-off. This leads to a consideration of the effects of maintaining work on those who carry it out. For the inmate researchers in particular, there were significant costs to their work. College programs ran counter to the survival strategies in which passivity was a key means to early exit. Participation also made inmates vulnerable to greater levels of surveillance: Inmate researchers, for example, were subject to books and diaries being searched and notes taken away. Moreover, exposure to education was at times painful for inmates because it disrupted established ways of thinking and behaving and brought into sharp focus the power differentials that characterized their lives, inside and out of prison. At the same time, involvement in and working to maintain the college program also had powerful positive effects: For the women at Bedford Hills, 80% of whom carry scars of childhood sexual abuse, biographies of mis-education, tough family and community backgrounds, long lists of social and personal betrayals, growing back the capacity to join a community, engage with a community, give back, and trust are remarkable social and psychological accomplishments. (Fine et al., 2004, p. 102) The implication is that the work of maintaining institutions may involve a complex juggling act of costs and benefits in which individuals find themselves having to cope with and manage complex social and psychological tensions. The inmates invested in college work were committed to education and its power to “liberate” but had to recognize that its message of freedom would be partial and that the work of maintaining access to the freeing power of education would also reinforce the restrictions under which they lived. Participatory Action Research and the Work of Disrupting Institutions We have argued that studies of institutional work and institutional research more generally has largely overlooked the issues of what motivates and mobilizes different interests to disrupt established practices and the skills and resources that enable actors to engage in institutional disruption, such as “cultural competence” (Lawrence & Suddaby, 2006, p. 238) or status (Battilana, 2006). An encounter with participatory action research could provide a way to explore the motivations and orientations associated with actors working to disrupt institutions. Participatory action research could also offer a way to explore the resources and skills associated with disrupting institutions. Existing institutional research has connected the disrupting of institutions to direct influence over control mechanisms, as with professional associations, regulatory agencies, and the courts (Dezalay & Garth, 1995; Greenwood, Suddaby, & Hinings, 2002; Jones, 2001). Less explored are the skills and resources available to those most disadvantaged by existing institutional arrangements—marginalized actors who have perhaps the greatest motivation to disrupt established practices. A powerful illustration of how and why actors might work to disrupt institutions comes from a participatory action research study situated in Guatemala following the 1996 Peace Accords (Lykes, Blanche, & Hamber, 2003). The project focused on “recording and critically analyzing multiple stories of daily living, that is, of war, its effects, and ongoing poverty.” A total of 20 local Maya Ixil women photographed life in their own and neighboring villages, taking pictures of women and their families. From the thousands of pictures and hundreds of stories collected, the PhotoVoice participants selected 60 photos and 11 stories “though which to re-story the massacres, displacement, death and destruction that characterized life among them during the war” (Lykes et al., 2003, p. 82). The following is a brief excerpt from one of those stories. There were 200 people massacred in 1982 in the village of the Finca La Estrella. We feel extremely resentful because of this terrible tragedy and we are saddest of all for these victims who were murdered so suddenly. They didn’t know that they were going to die because they were campesinos, [peasants] workers, who were guilty of nothing. They didn’t owe anything to anyone, but what is saddest is that the children and babies were murdered and they were only children, no more. (Women of PhotoVoice/ADMI & Lykes, 2000, p. 27) The work of this participatory action research project was disruptive in a number of ways. At a local level, the project disrupted the traditional relationship between Mayan people and photography, which has been dominated by tourists and professionals seeking to capture exotic customs. In contrast, the participants in this project became the first rural Ixil and K’iche’ women to become “professional” photographers, with computer skills and with the ability to balance financial accounts, write grant proposals, and speak at national forums. It disrupted the relationship between local people and social science by allowing peasants to appropriate “the skills and techniques of social scientific research in the service of speaking out about past horrors” as well as gender roles and other cultural practices through the introduction of cameras into the local economy. Most generally, the project has disrupted the scientific and human rights discourses about state-sponsored violence by injecting rural women’s voices, thus “transforming the ‘talk’ as well as the lives of those who speak their truths” (Lykes et al., 2003, p. 85). This participatory action research study provides significant insights into why and how actors, and especially marginalized actors, attempt to disrupt institutional arrangements. Examining the motivations of these actors highlights the importance of the local and the personal. Unlike relatively elite actors examined in other contexts (Greenwood & Suddaby, 2006; Sherer & Lee, 2002), these women were attempting to effect incremental shifts in the local village economy through the development of increased capacity. These motivations are reminiscent of those examined by Marti and Mair (2009) in their study of entrepreneurship in Bangladesh, in which they argue that the institutional work of the very poor is often distinctively experimental, provisional, and simultaneously transformative and enhancing of existing institutions. At the same time, the motivations of the Guatemalan women were deeply personal and political, connected to their experience of the war, in which they were “silenced by ongoing terror, gender relations in rural communities, local political and religious power dynamics, and pressing concerns for the material survival of themselves and their children” (Lykes et al., 2003, p. 81). Thus, the participatory action research study was for these women a means of giving voice to their experience and transforming the public record of the war, its meaning, and its effects on rural Guatemalan women. This study also sheds light on the issue of how marginalized actors might gain the resources and skills to disrupt institutions. In this study, several factors emerge as critical to this issue—collective organizing, connections to elites, developing a language for reflection and action, and remaining grounded in one’s own life experience. An important factor in understanding how the rural Guatemalan women were able to engage in disruptive action was their collective organization as the “Association of Maya Ixil Women-New Dawn.” Collective action has long been associated with the empowerment of marginalized actors (Alinsky, 1971; Hargrave & Van de Ven, 2009), facilitating both the consideration of more radical alternatives and greater risk taking. A second important factor highlighted by this study is the importance of connections to elites. Although the photographers were rural women with very little access to economic resources or political power, the success of the project was also dependent on the involvement of a professor from an elite U.S. university, who was able to furnish the project with the necessary hardware as well as training in both social science and photography practices. Critical to the success of this participatory action research study was the participants’ acquisition of a system of representation—a language—for critical reflection and action. This language was both visual and verbal, involving the use of photographs to document their conditions and the collection and analysis of women’s stories. These practices represented the construction of a shared cultural toolkit (Swidler, 1986) that allowed the participants to construct disruptive representations of their own realities. Finally, this study shows the importance of disruptive action being grounded in the personal experiences and emotions of the actors. Disruptive action is often difficult, risky, and slow and so can require a certain faith and patience that may be difficult to maintain if the task is not directly meaningful to the actors. In this case, the women were photographing and analyzing their own stories and the stories of women with whom they closely identified, focusing on the horrific experiences of a very recent war. Reflections: Photographs, Memories, and Plans for the Future All gap years come to an end, and the traveler is faced with some significant choices. Does he or she choose to see his or her experience as an interesting and stimulating but ultimately unique event? Or does he or she work to translate and integrate his or her discoveries into his or her “‘normal” life? We have argued that an encounter with participatory action research could offer the study of institutional work a way to address a set of issues that have been somewhat neglected within institutional research. We conclude this article by considering the relative merits of three stances toward the potential role of participation action research in the study of institutional work: a thought experiment, a research method to be leveraged, or an approach, the central tenets of which could be actively adopted.

Happy memories: A thought experiment. As far as we know, no studies have explicitly connected institutional work with participatory action research, and this may remain the case. It may be that institutional scholars will at best see connecting with participatory action research as the basis for useful thought experiments. Through such experiments, institutional researchers might explore protracted puzzles and challenges in a new light without needing to leave the comfort of their intellectual homes. Such an approach would not require insti­tutional scholars to actually do anything different beyond engaging in, potentially generative, cognitive play. But this would be a substantially missed opportunity. First, if we accept that organization studies, and institutional theory in particular, have become increasingly disconnected from real-world issues, then thought experiments alone are unlikely to foster the substantive investment needed to establish research agendas “grounded in the real world of organizations and people” (Lorsch, 2009, p. 111). Second, much of the value of participatory action research is gained only from participation—by engaging in the practices that connect academic and non-academic researchers and foster the development of novel insights and solutions. It is one thing to acknowledge the value of the lived experience of participants and how it could engender understanding of institutional dynamics, but it is quite another to actually see the relationship between lived experience and institutions in practice. This requires an approach that gets “up close and personal,” to observe and engage with the day-to-day realities of individuals and orga­nizations to understand how patterns of thinking and acting are created, maintained and disrupted. New tools: Participatory action research as method. An alter­native approach is to draw on the methodological practices of participatory action research and integrate them into institu­tional scholarship. Participatory action research has drawn on a wide variety of creative modes of engagement with partici­pants (e.g., Ospina, Dodge, Foldy, & Hofmann-Pinilla, 2007) such as storytelling, photography, poetry, drawing, sculpture, drama and popular theatre. Adopting more flexible research approaches and styles might facilitate creative collaborations among researchers from different disciplines and back­grounds, as has long been called for in our own publications (Polzer, Gulati, Khurana, & Tushman, 2009; Zald, 1993). But are there are real problems with this approach. First, research methods require influential academic arbiters to accept their legitimacy. Even with the substantial investments of time and energy by institutional researchers to acquire the neces­sary skills, these tools may not sit comfortably with existing methodologies. Methods associated with participatory action research emphasize the active involvement of participants in the research process, and in doing so, they can challenge well-established measures of research quality, such as researcher objectivity. Second, these new methodologies may lose their foundational and motivating roots in transla­tion. The eclectic use of research methods in participatory action research is motivated by a goal to assist individuals and organizations to affect social change. The research meth­ods in themselves are merely the means to reach a greater end. This is in contrast with mainstream academic practice where method selection is the preserve of “expert” research­ers and where demonstrating methodological mastery can be prized over results. Integration: Participatory action research as foundational. A third option for those seeking to connect to participatory action research is to actively integrate its foundational ele­ments into their work. At the end of a gap year, some return­ing travelers bring home not only photographs and memories but also new ways of thinking and acting in the world that they integrate into their home environments. We recognize that significant risks accompany an integrative approach. Objectivity is a highly prized quality in academic discourse. Empirical settings are described in terms of their potential for providing theoretical insight rather than for the chance to effect social change. Agnosticism with respect to the impacts of institutional arrangements, however “unjust or dysfunc­tional,” is the dominant orientation in published institutional accounts (Greenwood et al., 2008). Institutional scholars seeking meaningful connections with participatory action research may find themselves straining against the demands that accompany academic privilege or living in between two worlds, unable to reconcile the differences. Despite these risks, we believe that the potential gains from truly integrat­ing participatory action research into the study of institu­tional work would make the effort worthwhile. For institutional scholars seeking a substantive connec­tion would mean incorporating central ideas from participa­tory action research, including the importance of tacit, local knowledge, the coupledness of action and insight, and the role of research in social change. We believe that adopting these ideas could significantly affect the study of institu­tional work in a number of ways—the questions we ask, the goals of our studies, our relationship with our research “sub­jects,” the epistemological bases of our analyses, and what we mean by dissemination. Studies of institutional work might begin to focus on research questions developed in con­junction with members of the communities we study, with the aim of both developing theoretical insights and improv­ing the living conditions of participants. These shifts could lead to exciting new directions for institutional research that significantly enhances its impact on the world outside aca­deme and injects new life into its theoretical discourse. In this article, we have sought to explore an unlikely pairing and examine the potential value of taking the study of institutional work on a trip to encounter the strange world of those who want to change the world through participatory action research. We have tried to show the potentially deep, but unexplored, connections of institutional work research with participatory action research, as well as how it might be directly affected by a more extended encounter with the aims, methods, ideas, and practices of this change-oriented approach to research and action. What we found was a wealth of potential new insight and direction for the study of institutional work—changes that could make it more effective both as a source of scholarship and as a foundation for action. In truth, we do not expect all institutional scholars to “drop their tools” (Weick, 1996) —our intellectual and practical commitments tend to make us reluctant to accept radical epistemological or methodological shifts. But we do believe that for those researchers seriously interested in getting at the practices of institutional work, adopting some of the approaches and ideas of participatory action research could be of tremendous value. Moreover, for those of us who long for a more impactful institutional theory, we hope that engagement with the tenets of participatory action research might provide a path forward. Note 1. This problem is not specific to institutional theory, of course. A range of authors (Dess & Markoczy, 2008; Lorsch, 2009; McGahan, 2007; Pfeffer, 2007; Rynes, 2007; Shapiro, Kirkman & Courtney, 2007; Van de Ven & Johnson, 2006) have argued that management theory in general is problematically disconnected from practice. We focus on institutional theory here, not because of its distinctiveness, but because we believe that overcoming this disconnect will be achieved most effectively within academic subcommunities that share important assumptions about social reality and the practice of management.

tive of The Lilly Endowment, which made resources available to help increase the number of Hoosiers with college degrees. From the outset we recognized that graduation was one of a number of highly correlated indicators of student success that included student satisfaction, student effort, and a wide variety of desirable learning and personal development outcomes (Astin, 1993; Pascarella and Terenzini, 1991). For this reason we adopted a broad, holistic definition of student success that would resonate with a wide spectrum of institutional agents, in contrast to exclusively emphasizing retention activities. Many faculty members, for example, find the term “retention” off-putting and inconsistent with their perceptions of the appropriate goals of university study. Also, focusing on student success was consistent with strategic priorities articulated in university planning documents. For example, the Indiana University Bloomington campus plan (1995, p. 1) asserts that IUB “will place student learning, intellectual exploration, and educational persistence at the center of its missions.” PROMOTING STUDENT SUCCESS IN THE FIRST YEAR We decided to concentrate initially on the first year of college because it is during this time that the largest fraction of students leaves (Tinto, 1993). The literature on effective colleges, teaching and learning, and student development processes and outcomes suggested that four clusters of activities were critical to promoting student success in the first year (e.g., Astin, 1993; Barr and Tagg, 1995; Chickering and Gamson, 1987; Chickering and Reisser, 1993; Dill, 1999; Education Commission of the States, 1995; Ewell, 1997; Kuh, 1993; Kuh et al., 1991; Pascarella and Terenzini, 1991; The Study Group, 1984; Tinto, 1993; Upcraft, Gardner, and Associates, 1989). These are (1) designing appropriate anticipatory socialization experiences, (2) strengthening academic foundations, (3) integrating academic and social experiences, and (4) using a systemic collaborative approach to coordinate, implement, and institutionalize the interventions. The following sections provide selected examples from each of the four clusters that represent more than 2 dozen complementary initiatives. Anticipatory Socialization Anticipatory socialization is the process by which newcomers become familiar with the values, attitudes, norms, knowledge, and skills needed to perform in a new role or environment prior to actually entering the setting (Bragg, 1976). Students who have a clear understanding of what college life is like are more likely to make wise college enrollment decisions (Hossler, Schmit, and Vesper, 1998), devote higher levels of effort to educationally purposeful activities (Kuh et al., 1991; Pace, 1990), report higher levels of satisfaction than other students Illustrative Initiatives Universities can increase the chances that new students will successfully manage the transition to college by communicating to them clear, consistent messages about what they can expect academically and socially (Kuh et al., 1991; Pascarella and Terenzini, 1991). The months prior to matriculation and the first 6 to 8 weeks of the first semester are considered critical because it is during this period that students form impressions about the university environment and whether they “belong” in college (Upcraft et al., 1989; Pascarella and Terenzini, 1991; Tinto, 1993). Personnel from the IUB Office of Academic Affairs, College of Arts and Sciences, Enrollment Services (which includes admissions, financial aid and orientation), Campus Life Division, and Registrar began to meet regularly in the spring of 1997. The group constituted what management experts (e.g., Senge, 1990) call a cross-functional team and euphemistically labeled itself “Frosh Up” to connote its commitment to lifting up all new students to the level required for academic and social success. The group spent about 4 months discussing its charge, becoming familiar with the research on effective transition experiences and programs, and inventorying the extent to which the pre-college and early fall post-matriculation experiences of students new to the campus were consistent with the university’s educational goals and academic values and with the best practices described in the literature (Kuh et al., 1991; Tinto, 1993; Upcraft et al., 1989). Two primary concerns were noted: (1) large numbers of new students typically moved into campus housing several days prior to the beginning of formal orientation events, and (2) orientation events emphasized social activities more than academic preparation and intellectual life. As a result, the university had little chance to counter the potentially deleterious influence of certain aspects of contemporary student culture on attitudes and behavior patterns and to socialize students to acceptable academic norms. Committed to infusing best practices in new student recruitment, summer orientation, and fall welcome week, staff members in the Offices of Admissions and Student Financial Assistance undertook a systematic review of their activities and the messages they were sending to prospective and matriculating students about academic and social norms, expectations for campus life, and the behaviors associated with academic success. An effort was made to increase student-institution fit by communicating to students and their families the institution’s values and expectations, especially the importance of balancing the time and energy students devote to in-class and out-of-class activities. Orientation program staff then evaluated and redesigned their print materials and programs in order to reinforce the academic messages the admissions office was sending to new students. An academic expectations video filmed by currently enrolled students was given to every matriculant at summer orientation and students and their family members were strongly encouraged to watch it together. A follow up survey indicated that about two thirds of those who saw it (80% of the respondents) viewed it together as a family and about three fourths found it helpful in preparing for the transition to the university. Finally, personnel from Orientation, Enrollment Services, University Division, Academic Affairs, Registrar, and Residential Programs and Services worked together to redesign the fall welcome week so it would focus more on inculcating academic values. Toward that end they reduced the length of welcome week, increased the amount of academic and intellectual content in programs and events, and shortened the amount of time (from 6–9 days to 4 on average) students were on campus prior to the first day of classes. The shorter, more compact schedule substantially reduced the amount of unstructured free time students had and resulted in record attendance (7,000+) at the annual Freshman Induction Ceremony and the President’s Picnic (5,000+). Residence life staff report that first-year student participation in campus cultural events is on the rise. Academic Foundations Reaching one’s potential requires a strong foundation of academic skills and study habits. Doing well in introductory “gateway” classes is especially important because these courses tend to have large enrollments and students must pass them before they can begin taking major field classes. Particularly important at large universities is reducing the psychological size of the campus and individual classes in order to encourage greater levels of student involvement in educationally purposeful activities (Boyer Commission, 1997; Kuh et al., 1991). Institutional research at IUB showed that students were more likely to use tutoring and academic skills centers if they were accessible and available when most students wanted to use them, typically between 6:00 p.m. and midnight. Illustrative Initiatives Once again we turned to the best practices literature to discover what educationally effective colleges and universities do to channel students’ energies toward appropriate activities and engage them at a high level in these activities (Chickering and Reisser, 1993; Educational Commission of the States, 1996; Ewell, 1997; Kuh et al., 1991; Pascarella and Terenzini, 1991). Perhaps the best known set of engagement indicators is the “Seven Principles for Good Practice in Undergraduate Education” (Chickering and Gamson, 1987): student-faculty contact, cooperation among students, active learning, prompt feedback, time on task, high expectations, and respect for diverse talents and ways of learning. All are positively related to student satisfaction and achievement on a variety of dimensions (Astin, 1993; Goodsell, Maher, and Tinto, 1992; Sorcinelli, 1991). At a campus as big and organizationally complex as IUB it is not possible to immediately adopt and use good practices on a large scale. We concentrated on: (1) revising selected introductory courses to engage students at higher levels; (2) convening a workshop (the Freshman Learning Project) for faculty from different disciplines to share “what works” in teaching first-year students; and (3) helping underprepared students acquire requisite academic skills by providing critical literacy courses and making academic support services readily accessible. Illustrating some of the features of this cluster of activities is a coordinated set of interventions for a particularly challenging gateway mathematics course, M118 (Finite Math), which is required for a host of majors including business. Historically, about 38% of the students enrolled each term earned a D or F grade or withdrew from the class. The most radical aspect of this initiative was developing a 2-semester reduced- pace equivalent to M118. This option exceeded expectations in that failure and withdrawal rates dropped substantially with the vast majority (93%) of students in the 2-semester reduced pace sequence achieving grades of C− or better in the first semester. About 80% of those going on and taking the second semester of the course earned a C− or better. Moreover, the DFW rate for students in the reduced pace sequence was only 15%, compared with 65% in M118. Our institutional research showed that students in high-risk courses were almost twice as likely to seek tutoring when it was available in their own residence hall than when the same service was provided at other campus locations. Academic support centers (ASC) were established in two residence halls; the first in 1996 and the second in January 1998. Controlling for variables such as SAT scores, high school class rank, and math ability, ASC freshman users dropped fewer classes in the second semester, had higher first- and second semester grades, and persisted into the sophomore year at a higher rate than their counterparts who did not use the ASC (89.6% compared with 81.7%). Students who visited the centers regularly for mathematics help consistently earned higher grades in their mathematics courses than students of similar skill and ability levels who did not use these services. In fact, students visiting the centers at least 10 times for help with M118 earned grades that were on average 0.8 points higher than comparable students who did not, suggesting that a weekly visit to an ASC could potentially improve one’s performance by almost one full letter grade. Thus, the combination of the reduced-pace course plus access to study skills support boosted academic performance enough to help a nontrivial number of students break the cycle of withdrawal and failure that is costly both to the student and to the institution. Integrating Academic and Social Experiences When in-class and out-of-class experiences are mutually supportive, students gain more from their college experience (Kuh et al., 1991; Pascarella and Terenzini, 1991). Certain features of a residential university can be intentionally arranged to achieve this complementarity, thereby influencing to some degree the amount of effort students devote to various activities and what they learn. Illustrative Initiatives The goal of this set of initiatives was to integrate student academic and social experiences to a greater degree than might otherwise happen. Intellectual and cognitive development are not the exclusive province of so-called academic or in-class tasks, but are associated with a variety of social as well as intellectual experiences (Kuh, 1995; Pascarella and Terenzini, 1991; Terenzini, Springer, Pascarella, and Nora, 1995). Gains in cognitive complexity are a function of a combination of peer interactions, academic-based activities such as studying or talking with peers and faculty about academic matters or other issues related to their studies (e.g., advising, paper topics, graduate school), and leadership responsibilities (Pascarella and Terenzini, 1991). Living in a campus residence units organized around academic or intellectual themes is associated with a variety of desirable outcomes including higher-than-average gains in critical thinking, self-awareness, social competence, self-esteem, and autonomy (Kuh, 1995; Kuh, Douglas, Lund, and Ramin-Gyurnek, 1994; Pascarella, Terenzini, and Blimling, 1994). With this information in mind, four clusters of activities were pursued: (1) creating small, personalized academic communities (Freshman Interest Groups) to help students successfully navigate the academic transition to the university; (2) enhancing the campus living-learning environment by increasing the number of thematic housing units and academic focus floors; (3) emphasizing academic achievement as a key role of residence hall staff through changes in selection and training activities and staffing assignments, and (4) increasing the number of mentors for minority students. Freshman Interest Groups (FIGs) at IUB are clusters of up to 20 first-year students who co-enroll in three courses and take a FIG Seminar together; the majority of students in a FIG also live in the same residence hall. The seminar is a 1-credit course designed to help students with the academic demands of university study; it focuses on the acquisition of note taking, time management, and critical thinking skills and also introduces students to the vast intellectual and cultural resources of the university. Thus, participating in a FIG connects students with peers who have a shared intellectual experience in a common social setting and provides positive role models. The students joining FIGs had significantly lower high school ranks and average SAT scores than the freshman class in general. However, GPAs and persistence rates for FIG students did not differ significantly at year’s end from the overall freshman class. It appears that the FIG experience may have helped ameliorate some initial academic deficiencies, but that more work can still be done to assist students who enter the university with weaker academic skills. The number of students living in an educational theme unit increased about 58% over the past 3 years in response to the creation of more such units. The organizing themes range from the Collins Living-Learning Center (now in its 28th year) and honors floors to language houses (the most recent addition is Dutch and Yiddish), to Health Sciences; Legal Issues; Media; Religion, History, Ethics and Philosophy; and Visual Arts; to Volunteer Service; and Women and Society, Outdoor Adventure, International Affairs and Global Markets, and Environmental Concerns. In general, students living in thematic communities have higher GPAs compared with other residents, tend to persist to the second year at higher rates (93% compared with the overall rate of 84%), and devote significantly greater effort to educationally purposeful activities compared with students living in other units. Campus-Wide Implementation Structure and Collaboration Any large, organizationally complex research university faces special challenges when attempting to create conditions that foster student success. An ethos of collaboration must be cultivated among administrative and student support units and faculty, academic administrators, and student life professionals in order to blur the boundaries between academic affairs, student affairs, business affairs, and other functional units and to promote the cooperation needed to improve the undergraduate experience. Illustrative Initiatives This set of initiatives was designed to coordinate campus-wide efforts to promote student success. As alluded to earlier, focusing the effort of many individuals and groups requires knowledge of the entire institution, not just its parts (Senge, 1990), as “total quality” approaches emphasize (Seymour, 1992). Timing can be important, too (Bolman and Deal, 1991). Ideally, the process might unfold coterminously with the arrival of one or more supportive administrative officers or be stimulated by an infusion of energy and resources, such as in our instance with the Lilly Endowment grant. Open dialogue and collaboration are critical at all levels of the institution in order to appropriately and expeditiously adapt organizational arrangements and processes (Ewell, 1997; Kuh et al., 1991; Senge, 1990). Self and organizational knowledge increase as people learn more about the work of others and how their own work influences others and the institution as a whole (Senge, 1990; Seymour, 1992; Wheatley, 1992). Effective cross-functional dialogue is especially valuable as it legitimates open communication and generates insights that only collaboration can produce. Such dialogue also supports the necessary shift in culture from one characterized by hierarchical administrative behavior to one that values partnerships and cross-functional collaboration between academic and student affairs (Birnbaum, 1988; Kuh, 1996; Senge, 1990). Our initiatives included (1) a visit to another campus by a group of 16 people that helped to develop a common vision and language for the student success initiative; (2) the formation of a 17 member campus Retention Grant Steering Committee chaired by the chancellor; (3) periodic progress reports by the chancellor to various campus governance groups to symbolically support the individual initiatives and call attention to the cumulative effect of the interventions; (4) the formation and periodic meetings of Implementation Teams made up of people directly responsible for implementing the activities to insure accountability; (5) an annual large scale event; and (6) cross functional clusters charged with monitoring progress and modifying implementation strategies. Space limitations permit only a brief comment about two of these. To promote broad inclusive participation, it helps to expose people to the views of outside experts. One option is a large scale event focused on a specific topic of interest that involves large numbers of stakeholders (Brigham, 1996). Our event took the form of an annual day-long spring symposium jointly sponsored by academic affairs and student affairs. Both symbolizing and modeling the importance and power of collaboration, this event has to date featured such national experts as Jean MacGregor, Alexander Astin, and Marcia Baxter Magolda as “idea champions.” Annually more than 40 concurrent sessions led by faculty and staff members have focused on effective strategies for promoting student success. Frosh Up, the cross-functional cluster mentioned earlier, continues to meet and determine how to best deal with emerging issues related to student success and how to make the most effective use of campus resources to address these issues. Except for a core group of academic and student life administrators, its membership changes depending on the issue. For example, when the FIGs program was being developed, the registrar and academic advising units were key players. Now they participate only when relevant issues warrant their consultation and action. Through their primary institutional teaching and administrative assignments, Frosh Up committee members reinforce and connect certain Lilly related interventions to other campus efforts to improve undergraduate education. One such example is integrating IUB’s Scholarship of Teaching and Learning (SOTL) initiative sponsored by The Carnegie Foundation for the Advancement of Teaching into the program of the annual spring symposium. Specifically, a dedicated concurrent sessions program track was part of the April 2000 symposium featuring faculty members involved in SOTL presenting various aspects of their work. DISCUSSION AND IMPLICATIONS Taken together, our efforts to improve the undergraduate experience at IUB are characterized by an emerging campus ethos marked by a commitment to student success guided by principles of good practice distilled from the higher education literature and research. The evidence to date suggests that this set of complementary research-based activities coupled with other programs are having the desired effects. For example, the total number of sophomores in fall 1999 was at an all-time high of 7,257. Eighty-four percent of the 1998 freshman cohort returned to the Bloomington campus for the 1999–2000 academic year, a first-to-second year persistence rate comparable to 1997, 1% higher than in 1995 and 1996, and about 4% higher than in 1994. This slight increase was realized in the context of record numbers of freshmen the past 2 years as we noted in the first installment of this article. Another important indicator is the improved persistence rates of minority students that are now approaching parity with the overall campus average. All this occurred without any significant change in the characteristics of matriculating students as we explained previously. These positive results have reinforced the common sense of purpose we labored hard to establish and has buoyed our confidence to press on with additional reforms. Some may conclude that the circumstances at Indiana University were unique and that attempts to use the higher education literature to guide similar efforts at other institutions will not bear fruit. While the situation at IUB may have been unusual (as are most campuses most of the time!), it was not unique. Many colleges and universities have obtained support to improve retention either from external sources or through reallocation of institutional funds. The Indiana story is instructive because of the way we intentionally used the higher education literature to guide the substantive elements of our interventions and the processes we employed to move the interventions forward. Although a few key interventions were underwritten by Lilly monies (FIGs, certain mathematics interventions), other equally important activities—such as the revision of admissions materials, the revamped summer orientation and fall welcome weeks, and development of additional theme-based housing units—had little or no external funding. Certain faculty development activities that promise to positively affect teaching and learning on campus for many years, such as the aforementioned Freshman Learning Project, were funded internally as part of a larger university system-wide set of quality improvement efforts. Our sense is that aside from supporting some initiatives, the most powerful influence of the external funding was that it helped focus institutional effort and made promoting student success a worthwhile, legitimate goal. The institutional improvement lesson is that creating a sense of urgency for doing something important and meaningful that is consistent with the university’s educational mission helps to focus effort and encourage collaboration on the part of key personnel from various corners of the campus. Under the umbrella of promoting student success we brought people together in order to coordinate and increase the complementarity of different programs, some which had been going on for years without external support. Thus, we believe the increase in student persistence the campus has realized is due in large part to linking various initiatives so that the sum of their impact could be greater than what the individual projects could produce. To expand the number of people committed to the effort we brought small groups together to infuse promising practices from the literature or in use at other schools that would have direct implications for their work. For example, as previously mentioned, we arranged for a core group of academic administrators, faculty members, and other staff to visit another residential research university campus to see and hear firsthand from people like them about the promising educational practices being implemented at the sister school. This was a very powerful step in making the research and theory that undergirded their work accessible and useful. Moreover, the trip introduced the participants to the vision and language of student success in ways that no series of meetings on our campus could, convincing key people that adopting promising practices was not only doable but preferable to the way the campus was currently operating. Consistent with the advice of modern management theorists (Bolman and Deal, 1991; Schein, 1993; Senge, 1990; Wheatley, 1992), we opted for being organic, intuitive, but always intentional and sensitive to cultural and political issues. We acted rather than spend countless hours carefully crafting a elaborate blueprint of preferred events and programs or seeking approval to experiment. Because our interventions were guided by best practices from the literature, we were comfortable with pressing ahead with implementation, evaluating the impact of the respective activities as we went. We shared information widely about the status of interventions and their effects on students with various groups of people (faculty, student life staff) at various levels (academic and student affairs units, campus-level offices) so that they could better understand the implications for student and institutional performance. We studied and refined evaluation data before reallocating human and other resources. Short-term successes were applauded, but the focus was always on the larger, long-term goal of improving the quality of the undergraduate experience for all students. Occasionally we encountered the unexpected, such as a proposed change in the eligibility requirements for students seeking readmission, a policy with the potential effect of decreasing campus retention figures in the short term. However, after a spirited debate in the Frosh Up group we realized that the likely long term implications were consistent with our overall goal: helping individual students obtain the academic skills and commitment needed to succeed at the university. Serendipity happens on any campus. Whether institutions improve as a result depends on whether governing board decisions, retirements, new hires, student preferences, and other exigencies of institutional life will be seized as opportunities or overlooked. In the first part of this article we described the creation of the enrollment services division. In the early 1990s, the responsibility for orientation programs was transferred to the Office of the Vice-Chancellor of Academic Affairs. In another independent administrative move, the formerly separate functions of housing (physical plant, maintenance, food service) and residence life (educational programming, advising) operations were merged in 1998. The directors of both the newly formed Enrollment Services and Residential Programs and Services operations were professionals conversant with student development research and institutional improvement strategies; moreover, they were committed to collaboration. These changes made cooperative ventures such as implementing FIGs and emphasizing academics in the residence halls easier to implement, but they did not necessarily ensure that this would happen. Without question, it is a great advantage to have like-minded people knowledgeable about the higher education literature committed to student success and the implementation of complementary programs. Creating a common vision and speaking a common language speeds up the collaborative process and makes it easier to recruit and teach others about the task at hand. We intentionally arranged work groups and agenda-building sessions to regularize contact between key leaders and their associates with an eye toward creating an organizational synergy that would not have occurred without such structuring of cross-functional groupings. As a result, the research and theory referenced in this article became accessible to much larger numbers of people than might have happened otherwise. This information is continuing to influence institutional policy and decision making in many unintended and unanticipated ways. CONCLUSION There is plenty of fruit on the vines of higher education research to nurture and sustain efforts to improve the quality of the undergraduate experience. We used the literature primarily to guide the development, implementation, and articulation of mutually reinforcing interventions. We supplemented the research literature with an ongoing program of institutional research and evaluation to provide real-time data to guide our efforts. There is, of course, much more to this unfolding story. Other interpretations will come to the fore of why and the extent to which these efforts seem to have had a positive effect. Certainly the availability of external resources as an instrument of change was important, although there are numerous examples—both current and past—of the failure of grants to change important aspects of how colleges function. In the final analysis we believe that whatever these interventions have contributed is a function of the combination of being familiar with the relevant research (including its limitations), a focused program of institutional research and evaluation, and a will to act, all driven by a commitment to making a positive difference in the lives of individual students and the institution.

A Sense of the Problem As a result of these concerns, and others, IRBs have now been granted a mandate to oversee research processes more broadly, including the training of graduate students in qualitative research skills (e.g., interviewing, observation) and classroom-based exercises in such skill building. This situation is far less flexible than in the past when graduate faculty could submit syllabi when and if they were changed but otherwise, operated under more or less “blanket” approval for graduate-level training so long as the IRB was apprised of their intent to act as principal investigators for all class members. Thus, approval for conducting such exercises as part of a course requirement and course grade might be granted for a decade at a time. Heightened concerns regarding human subjects have changed all that. At the same time, pressure from the political right has intensified to discredit the products of postmodern theorizing, including constructivist theories of knowledge, postmodern epistemologies, Foucauldian analyses, post structural investigations, action and participatory action research, and other kinds of research associated frequently or primarily with qualitative and/or interpretive research (Bauerlein, 2001; Feagin, 1999; Koertge, 1994; Lincoln & Cannella, 2002). Between criticisms from detractors of certain theoretical approaches and a heightened sense of legal issues regarding medical protection (see, for instance, the recent halt of all federally financed medical studies at Johns Hopkins as a result of the death of a woman participating in medical research), the stances of IRBs have shifted from assuring that human subjects’ rights are protected toward monitoring, censuring, and outright disapproval of projects that use qualitative research, phenomenological approaches, and other alternative frameworks for knowing and knowledge. Some IRBs are quite clear and aboveboard that their main concern is protection of the institution from damage. This is a fundamental shift from the original purpose of ascertaining risk to humanand animal subjects and assuring that informed consent was adequate to prepare human subjects for associated risks. The American Association of University Professors’ (AAUP) (2001) “Protecting Human Beings: Institutional Review Boards and Social Science Research” report echoes some of these same issues. Three contexts in which this trend has been especially noticeable have emerged: externally 220 QUALITATIVE INQUIRY / April 2004 Downloaded from funded projects, student dissertation research, and qualitative research methods courses taught for graduate students. Data During the past decade, three forms of research and teaching activities have gone through IRB approval and disapproval. As IRBs increase their regulatory functions, the number of “stories” of researcher experiences is increasing. As the stories increase, the sense of frustration, anxiety, and anger appears to increase correspondingly. By way of informal and formal processes, the data collected for this particular piece of research exhibit high correspondence with testimony provided by social scientists and reviewed by the AAUP (2001) prior to preparing its report. Data were drawn from author experiences and from reports, letters, and interviews with other researchers at Research Extensive universities. The first context: Funded projects. In contradistinction to the policies of even a decade ago, when funded social science or educational research projects were assumed to have already proceeded through several levels of review and so presented little, if any, risk to research participants, scholars around the country are having some difficulty, with repeated efforts, getting already funded qualitative studies approved and through the IRB review process. It is not that the risks of most social science research have gone up; rather, it is that the scrutiny of such projects has increased exponentially. The issue of risk to research participants is not trivial. But as the AAUP (2001) pointed out, “This is not to suggest that risk-benefit analysis is inapplicable to social science research, but rather to emphasize a simple proposition: that different kinds of risks and benefits are associated with different kinds of research” (p. 61). Medical experimentation runs rather different kinds of risks to patients; in such research, participants can and have died. Social research, however, poses different kinds of risks. Research participants may be embarrassed, humiliated, feel their dignity has been compromised, or experience very real invasions and violations of their privacy. They may feel that they have lost control over some portion of their lives (a loss of agency) or that they have been made fools of (as occasionally happens when participants have been deceived regarding the purposes of research, which deception is approved under certain circumstances by federal law). These are not inconsequential concerns residing in human subjects protection, but it is clear that as the AAUP report parsed the issue, different kinds of research pose different kinds of risks; losing one’s life appears to be rather more serious than losing one’s dignity—although the individual research participants may not hold that view at the moment. Taken in the light of campus concerns for the importance of externally funded research and development projects, the stance of some IRBs toward more interpretive social science projects would seem strange. In view of concerns surrounding the IRBs and academic freedom, the somewhat ambiguous powers granted to these boards, and because of recent attacks on qualitative research, it is slightly less bewildering or puzzling. The second context: Student research in dissertations. Traditionally, unless research procedures seemed to indicate close supervision to assure the protection of research participants, student dissertations were frequently remanded to the “Exempt” category—that is, highly unlikely to require more than cursory review and unlikely to cause any damage or psychological harm to individual human subjects (even as they are equally unlikely to do any good)—and given review by a subcommittee rather than the full IRB committee. This process was swift, expeditious, and thorough, even though completed by fewer IRB committee members. At this point in time, reviews are taking far longer than the usual 6 weeks; dissertation work that is qualitative is undergoing full-committee review; and at some institutions, qualitative, phenomenological, critical theorist, feminist, action research, and participaory action research projects have been summarily rejected as “unscientific,” “ungeneralizable,” and/or inadequately theorized (even though they maybe descriptive, historical, or exploratory projects and therefore, unable to be theorized at the moment).Avariety of strategies have been devised by researchers to overcome persistent rejection by IRBs, including several that actually undermine the work but that have the effect of permitting graduate students to complete their doctorates (AAUP, 2001, p. 64; Confidential, personal communication, February, 2001). The third context: Qualitative research methods courses taught for graduate students. Although it is time consuming to submit IRB proposals for each course taught, it has not been an issue until recently when full IRB committee approval began to be required (rather than having the proposal go through “Exempt Status” approval procedures). Now, frequently more than 3 months is required to hear back from the IRB. And for the first author of this article, an advanced fieldwork methods course was approved, gotten under way, and at the midpoint of the semester, the IRB decided that the course syllabus should be rereviewed, and the entire course was subsequently disapproved with a warning letter that the course must stop as of March 7! Only the intervention of a dean, and a reminder that the university had moral, fiscal, and ethical, as well as contractual, responsibilities to students who had signed up for the course kept the course from becoming an intellectual “lockout.” The issues—reasons frequently cited as bases for rejection—seem to be nonquantitative or experimental research methods (i.e., qualitative methods), new paradigms for inquiry (e.g., phenomenological, feminist, postmodern, Foucauldian, and/or constructivist), and lack of fit with traditional rigor criteria (e.g., generalizability, replicability, objectivity). Such reasons exhibit, at least on their face, either an unfamiliarity with nonquantitative methods for data collection and with postmodern and critical epistemologies or resistance to nontraditional, non–“scientific model” research methods and models. Or there is, perhaps, something far more ominous occurring: a backlash against qualitative research in all its non rationalistic or postmodern forms (Lincoln & Cannella, 2002; National Research Council, 2002; No Child Left Behind Act of 2001). Significance of the Problem The issue is critical. There are at least five grounds on which to be concerned for the future of qualitative research as a legitimate scientific inquiry mode. First, such widespread rejection of alternative research forms and ways of knowing suggest that qualitative research will be heard less in the policy forums in Washington and around the country in state legislatures. Second, the situation suggests that qualitative researchers are having to confront the control of “important discourse and decisions” by traditional elites rather than taking part in “open[ing] up that dialogue and decision making to the larger population” (Feagin, 1999). Indeed, among the AAUP (2001) committee’s conclusions was that some IRBs “too often mistakenly apply standards of clinical and biomedical research to social science research, to the detriment of the latter” (pp. 55-56). Third, it suggests that new, young researchers, trained in alternative epistemologies and research methods, will find their inquiries rejected before they even begin careers in educational research. Fourth, it is clear from government documents relating to the IRB and review process (National Bioethics Advisory Commission, 2001a, 2001b) that the working definition of research has not changed in more than a quarter of a century, despite wide debate and a plethora of new scholarship that suggests that conventional definitions are far too narrow and unnecessarily limiting of the inquiries of serious scholars. And finally, this resistance and rejection suggests that traditional researchers have already understood the power and compelling quality of qualitative data and have rejected its strong, “data near” claims to validity in favor of a more distanced social and educational research where “social scientists . . . have lost touch with the moral and practical concerns from which . . . [the] field emanated” (Feagin, 1999). This experience is even more widespread than the AAUP and the authors have found it to be. New national and federal policy and legislative initiatives are not only inimical to qualitative research, in some instances, it will simply not be funded with federal funds (e.g., the No Child Left Behind Act of 2001; see, e.g., House, in press). The implications for both educational researchers and their professional organizations (as well as funding agencies) suggest a coming crisis. Failure to obtain permission to conduct qualitative studies or mandates that such studies be conducted in positivist fashion will greatly undermine educational researchers’ ability to uncover hidden aspects of social arrangements that contribute to unequal schooling, lower persistence rates of minority college students, or other less transparent educational processes. Dialogue seems critical; at the same time, intense dialogue on some campuses has proven less than useful. But a clear understanding of the problem and the variety of contexts in which it operates is a strong beginning for coping with rejection of alternative models of research. Some flavor for the kinds of problems will be evident in the following descriptions of actual cases (names and institutions have been changed to protect students, researchers, and institutions; the cases are written in the first-person voice of the scholar providing the authors with the write-up). Case Study 1: Deanna Holcomb This is the case that provided me with the opportunity to speak my concerns to the IRB. Deanna, principal of [an] elementary school in [a nearby “bedroom community” to a large urban area], was doing a fairly innocuous study of her own school to obtain the perceptions of teachers, students, and parents regarding the first year of operation of a 4th grade teaming program. The purpose was the traditional one of a program evaluation: to observe how various stakeholders responded to an innovation and to use that input to make decisions regarding its continuation, abandonment, or modification. The data collection phase of the study was designed to cover a spring semester and the following school year. For nearly a full year the IRB and Deanna interacted on her study. The pattern was that they would require that she make additional changes, she would make the changes, and then they would request additional ones. As the end of the first spring semester, Deanna wanted to obtain some perceptions of parents before the study began. Since this part of the study was not being questioned by the IRB, she and I decided that, as part of her regular program evaluation activities, she could go ahead and get this information. The IRB got word of this and notified her that she shouldn’t have done this without their approval and that she could not use the data collected in her record of study. (Her advisory committee agreed that this would not destroy her study; but it had a chilling effect on her.) Finally, after all this tinkering with the proposal and IRB form, [the chairman] of the IRB invited me to make a presentation to that group. It was clear by that point (it hadn’t been before) that what really bothered the IRB was the fact that she was doing the study in her own school. I made a presentation explaining why this type of research was so important for principals to do in their own schools and provided examples how this same type of research was being done in a number of prestigious universities. But they didn’t budge. What bothered them most at this point was the fact that the responses to the open-ended questions were being turned into Deanna’s secretary, who then separated them from names and other identifying marks before she turned the data over to Deanna. (This arrangement had been made in response to earlier concerns about confidentiality.) Their reasoning was that the boss-secretary relationship was so close as to make confidentiality impossible. As a result, Deanna proposed that they be sent to an administrator at the district office who would do the same service for her. (This administrator was a close friend of Deanna’s—who else would do it?) This arrangement was satisfactory to the IRB, and Deanna proceeded with her study. She graduated in December 2000. Perhaps the most interesting note from the entire experience was a piece of my exchange with the IRB. The IRB members agreed that principals need to perform action research in their own schools, but that it was quite another thing for them to do it with the blessing of [this state university], as implied by the Record of Study or Dissertation. At this point, I thought I saw a fallacy in their argument, and I gleefully pressed my point: “In other words, since it’s okay (and commendable) for a principal to do research in her own school, what you’re saying is that you’re not really worried about human subjects per se, but rather about the potential impact on [this university].” My elation quickly evaporated when, without the least bit of hesitation or shame, they agreed with me. Case Study 2: Roberta Goodwin This case represents advisor learning; the [Deanna Holcomb] case was not entirely in vain. When Roberta and I talked about her doctoral study, her school had just received a High-Schools-That-Work grant. In order to provide an ongoing evaluation, I proposed that she use her leadership team for the project as a “research team” to provide useful information about the direction for the project and to facilitate their decisions in coordinating the project. When Roberta got ready to write her proposal, her study was simply one of accessing this existing database and, after augmenting with other archival data from the school, analyzing it and writing a splendid record of study. Case Study 3: Colleen McCormick This case is of interest, partially because it dealt with the most sensitive human relations of all the cases, but also because it underlines the IRB’s bottom line: protect the university. Colleen’s study sought to explore how direct feedback to teachers on their performance in the classroom by sixth grade students might be used for professional development. All the teachers were volunteers, and the entire process was guided by Colleen, who was an assistant principal in the school, though not the evaluating supervisor of any of the teachers in the study. Oxford County School District (traditionally a very conservative institution regarding risk in research) recognized the great potential benefit of the study and, after carefully reviewing it, gave the study its approval. However, it took several months and repeated modifications and guarantees before the IRB reluctantly approved the proposal. Colleen nearly abandoned the project in favor of a safe and sterile study. Fortunately she persisted and produced an exemplary study. Case Study 4: Charles Jacobsen Charles Jacobsen sought permission to conduct a fairly innocuous study that called for interviews with principals and African American teachers to describe the extent to which their leadership talents were put to use in their schools. However, he got caught in a series of minor changes in his proposal that were required by the IRB. He would dutifully make these—only to have that body identify the need for new changes. Finally, after he had made all the changes the IRB requested, that body changed their forms without notifying him, and the approval of his proposal was set back another month. Charles’s patience and persistence in the whole process were exemplary. His proposal was finally approved, and he is now working on his study. Case Study 5: Laurie Thompson Laurie Thompson is an assistant professor in the educational psychology department, nearing the time when she must go up for promotion and tenure. After having received several small grants, on which the work was done in a timely and exemplary fashion, she received a rather large grant (well over $200K) to perform a substantial research study on her area of interest: children and technology. She has received IRB approval from three school districts in the areas surrounding our university for her study, which seeks to understand how students interact with technology, explore its uses for school projects, and their own sense of self-efficacy when mastering various aspects of Web-surfing, research, and communication. The IRB has sent her proposal back no fewer than seven times, asking for increasing clarification on precisely what questions she will ask the 4th, 6th, and 7th graders, and whether or not she will “vary” from those questions. Furthermore, the IRB has asked her to explain how, with such a small sample (they had in mind, they said, around 500 school districts) she would be able to “generalize” from a mere 50 children. They have asked, in effect, that she “swear” that she will not probe for amplification, clarifications, extensions or other additional comments on the children’s responses. As of this writing, Laurie has had the contract for over 15 months and has not been able to observe or interview a single child. It seems likely that Laurie will receive no further grants or contracts since she seems unable to proceed with this one. Case Study 6: Dianne Flowers Dianne Flowers is the only student in this collection of case studies who was actually prevented from doing the study she envisioned. Ms. Flowers works in an alternative secondary school and has extremely good rapport with the students in that setting. She wanted to tell the Alternative School story from the students’ viewpoint, including extended case studies of several prototypical students. However, after many roadblocks, she agreed to simply collect written information anonymously from the students, and, together with teacher interviews and a review of student records, she will attempt to write “their” story. The study is considerably weaker than originally envisioned; but given the difficulty of convincing the IRB, I supported her making this compromise. Case Study 7: First Author I was teaching an Advanced Fieldwork Methods class, 2nd semester (Spring), having received IRB approval for students to engage in various fieldwork execises under my supervision. On March 15, I received a letter from the IRB, dated March 7, which stated that approval had been withdrawn, and that I was to immediately cease teaching the course (some 8 weeks into the course, or fully half a semester).When I called to inquire regarding why the course, having been approved, was now disapproved, I was given two reasons: the form I had used was not “current” this semester, and I must reorganize the information, and second, there was a possibility that students might be engaging in research with “protected” populations (e.g., students, prisoners, medical patients, etc.). I pointed out that I had drawn the IRB forms off the university’s IRB website, and was informed that the form had changed in the last month, and that I was to resubmit the form. I then pointed out that the syllabus specifically forbid any research to be undertaken with protected populations, in writing. The chair of the IRB Committee, however, was adamant: I must cease teaching this course immediately. I then took the problem to the Dean, who presumably went up the chain of command, but ended up informing the IRB Chairman that I had received approval for the course, had filled out the appropriate forms more than 12weeks in advance of the start of the course, and that the University had a contractual agreement with students to continue courses in which students were enrolled, unless there was some compelling reason to terminate a course. There was much other conversation regarding this course, but the upshot was that I was permitted to complete the course with the students. It was an example, I believe, of the peremptory quality of some IRB processes— to first approve, and then pull approval in the midst of an ongoing doctoral seminar in qualitative research. \*\*\* These cases illustrate, we believe, some of the difficulties that teachers, researchers, dissertation advisees, and scholars are currently having with their own IRBs. In the case of the students, there appears to be a growing—and inappropriate, from our perspective—concern with research in one’s own context. This is especially surprising considering the extensive literature that commends ongoing research as a mark of the “reflective practitioners,” to improve schooling practices, and in light of the developing methodologies for having teachers and principals engage in systematic and disciplined inquiry about their own professional practices. Another difficulty that seems to be appearing in some of the cases is the reluctance of IRBs to approve research with children, in some cases, despite school board approval and encouragement of such research. We have, however, little hope of understanding what schooling and learning mean to children, or how children view learning processes, or what processes keep children as learners engaged unless we can do careful and thoughtful research with and among children, and with their input and assistance, in the learning context. Yet a third difficulty is the problem of action research, participatory action research, cooperative inquiry, and other forms of community-engaged research. IRBs appear to be having considerable difficulty with either under standing or with supporting such research, even though action research models (whatever their particular emphasis) show great promise of involving stakeholders at the research site in meaningful dialogue on their own, indigenous, contextually determined needs. A fourth arena of concern appears to be the reasonable and realistic assessment of potential benefits versus potential risks. Clearly, in the studies above, we have some concern with the “regulatory” aspects of research risk assessment (Pritchard, in press) and very little concern for the potential benefits of the research enterprise. The National Bioethics Advisory Commission’s (2001b) report very clearly states that even within areas of research that need oversight, many individual studies will involve little or no risk to participants. Although current federal policies allow for some distinction between research involving minimal risk and research involving more than minimal risk, the distinction operates mostly in terms of how the research will be reviewed—that is, how procedures are to be followed. But the distinction should be based on how the research is pursued, how the participants are treated, and how the work is monitored over time. (p. 8, italics added) In all but one of the cases cited above, the level of monitoring is extremely high (including oversight by both dissertation committees, and school district IRBs and/or graduate faculty) and the participants are given multiple levels of protection, although they are simply providing information and data that they would be required to provide in the normal course of their professional duties. In addition, the question of how the participants are treated seems to be virtually nonexistent, because in several of the cases, the participants are engaged in the data collection effort as a part of their ongoing professional responsibilities and experiences. The problematic question remaining is how the research was pursued. In every instance, it was qualitative and some variety of ethnographic, action, or participatory research. The fifth area of concern appears to be that at least some IRBs are, by the admissions of their own members, more concerned with protecting the institution in which they work than in facilitating research or assuring human subjects protection. In no instance in the above cases was there any significant risk to the individuals involved as participants. In all the instances in which public school sites were involved, the research itself had already received prior approval by the districts’ own IRBs, which presumably means that the school districts at least share equally some of the monitoring and supervisory responsibility for assuring protection, privacy, and confidentiality with the companion universities. The concern with protection of the sponsoring university is certainly serious; but in no way should a concern for the institution’s reputation or risk be the first concern of the IRB. Rather, IRBs are, by legislative intent, constituted to assure the protections first and foremost of human subjects. It is by doing the latter that the former purpose is served, not the reverse. WORK THAT NEEDS TO BE DONE Clearly, there is work that needs to be done. The recent hearings conducted by the National Bioethics Advisory Commission and the National Science Foundation (AAUP, 2001) have gone far toward shedding some light on the set of problems that exist, particularly for social scientists (and concomitantly, educational researchers) who maybe doing research that involves minimal, if any, risk to human research participants. The AAUP(2001) report, in particular, stresses the question of “level of risk,” as do sections of the National Bioethics Advisory Commission report (2001a, see particularly Vol. I, pp. 74-80, 2001b, p. 8). Level of risk is one of two critical issues (the other is the definition of research, which we shall take up in a moment) that need further exploration and new guidelines. Dialogue might increase the possibility of enlarging the view of IRBs on criteria for approval and oversight. Level of risk and potential direct benefits. Weijer (1999, 2001) proposed a framework for the analyses of risks versus potential benefits that we believe should be used more widely and explored more vigorously on campuses. His analyses focus on what he termed “component analysis,” or separate analyses of which parts of the research offer the possibility of “direct benefit to research participants,” whereas other parts of the analysis of level of risk would examine elements of the proposal that have “the sole intent of answering the research question(s)” (National Bioethics Advisory Commission, 2001a, Vol. I, p. 76). The ethical intent is to question which research procedures offer “direct benefit” to research subjects (or their communities) and which offer only to answer the research questions. The National Bioethics Advisory Commission (2001a) defined this component-based approach as one that requires IRBs to sort research study procedures into these two types of components to determine their ethical acceptability. The first type consists of those components containing particular procedures that may offer the prospect of direct benefit to participants. The second type includes procedures that do not . . . their sole intent is to answer the research question(s). (p. 76) The recommendation of the National Bioethics Advisory Commission (2001a) is quite straightforward: To the extent possible, IRBs should independently weigh the risks and potential benefits of each type of procedure. The risks associated with individual procedures offering the prospect of direct benefits are justified in relation to their potential to benefit the participant in addition to their potential to generate knowledge, and those procedures designed solely to answer the research question( s) are justified in relation to their potential to generate knowledge. (p. 77) In layman’s terms, two assessments of risk and benefit must occur: that of the possibility of direct benefit and that of the potential of a given study to answer the research question(s). The assessment of risk lies in weighing the possibility of a given benefit to participants against the ability to generate knowledge and weighing the potential of the study to generate knowledge, per se. The ethical parsing represented by component analysis forces judges of a given piece of research (in this case, IRBs) to look not only at the possibility of a given study for generating knowledge but also at that possibility alongside and in tandem with the potential for direct benefits. This is critical, we believe, in at least six of the seven case studies reported on earlier. In the first six case studies, strong potential exists for direct and indirect benefits to participants (and to subsequent students and staff like them), whereas there is little attendant risk to participants. In four of the six instances, participants would be engaging in the same activities with or without the researcher’s interest in completing a dissertation; the dissertation merely makes the interaction with the IRB necessary while at the same time permits wider dissemination of the results of the study to other professionals who might benefit from its insights and from the knowledge produced. Thus, on criteria proposed by the National Bioethics Advisory Commission itself, the ongoing disapproval of the cases does not serve either ethical criteria or protective criteria. Rather, disapproval (or endless requests for changes, alterations, etc. in the research design, which often act to connote future disapproval or at least to discourage researchers) signals what one IRB group frankly admitted to one dissertation advisor: The interests of the institution (whatever they are) are more important than the interests of fostering sound research—even when the research is deemed important to the IRB itself. Definition of research. Another problem that appears to be emerging is in the definition of research now widely adopted by many IRBs (and now codified in the National Research Council’s [2002] report, Scientific Research in Education). The Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979) and its definition have served as the basis for the current definition in use throughout the clinical and social sciences. The current proposed definition (following Belmont closely) is the “regulatory definition of research in 45 CFR 46.102 [which is]: ‘Research means a systematic investigation, including research development, testing and evaluation, design to develop or contribute to generalizable knowledge’” (National Bioethics Advisory Commission, 2001a, Vol. I, p. 36). The National Bioethics Advisory Commission (2001a) itself, however, identified several problems with this definition, two of which are critical for the cases described above. The first is that the definition does not “include the important distinguishing concept of who benefits from the activity”—that is, it does not include a more expanded consideration of stakeholders—and problems with the “use of the term generalizable” (National Bioethics Advisory Commission, 2001a, Vol. I, p. 35-36). The long-term implications of this definition include the likelihood that many, if not most or all, IRBs are unlikely to consider potential direct (or indirect) benefits as weighty ethical issues when considering approval of research projects as well as the likelihood that generalizability— long a criterion of conventional scientific research—will continue to be a point of contention for studies where generalizability is neither sought nor desired. Thus, studies in schools—where the long-term goal of the study is to understand more thoroughly best practices, or the meaning-making activities of those engaged in schooling, or the experiences of children in schooling— are likely to fail to win approval on both important criteria: potential direct benefit and generalizability. More important, however, is the issue of what kinds of research will be permitted, given that generalizability is a definitional and standard criterion for “research.” It is here that qualitative research, in particular, in all its forms, appears to be losing the battle for IRB approval and sanction. Most, if not all, qualitative research, but particularly that which is phenomenological in orientation, constructivist in focus, or located within the “action” traditions (e.g., action research, participatory action research, cooperative inquiry, and the like) never has as its focus the criterion of generalizability. In fact, most of the models of inquiry listed above specifically eschew generalizability as inimical to other overriding philosophical tenets that make generalizability impossible to achieve (or a useless phantom to be chased).1 This is likely why the chairman of one IRB can report that he thinks such research is important but that it should not be done under the aegis of his university. So long as the regulations state that generalizability is a part of the definition of research, qualitative research studies will continue to suffer inordinate review and sometimes inappropriate revision, if not outright rejection. Furthermore, so long as potential direct benefits are not taken into account, studies in a researcher’s own context are likely to fail the approval process, even when such studies are the only means by which professional practice in education is likely to see improvement or revision. Thus, what is coming to be understood more widely in the social science community as inquiry philosophies that lead to deeper understandings of the lived experience, is being systematically rejected or subjected to inordinate and onerous review processes that undermine timeliness. Level of risk implies vastly different meanings between biomedical and/or psychological research and research on educational processes, but in fact, all these forms of research are being subjected to the same kinds of review without any adequate assessment of the actual potential for risk or harm to research participants. THE IMPLICATIONS FOR QUALITATIVE RESEARCHERS There are several implications for qualitative researchers in this IRB quandary. First, qualitative research is likely to take far longer to get under way simply because the institutional review processes may be unable to sort out level of risk appropriately. Second, it is likely to be looked at askance because it is conducted in the context in which the researcher works and is therefore believed to be an unwarranted risk to privacy or to freedom to consent without coercion or to withdraw freely on the part of research participants. Third, qualitative research may not be assessed adequately with respect to the proposed criterion of direct benefits (Weijer, 2001) to participants. This may be especially so in the case of the professional practice of education, in school improvement, or in understanding learning processes from the perspective of learners (children in particular). It may be easier for IRBs to assess the direct benefits to research participants when the research is biomedical (palliatives for pain, improved mental functioning for Alzheimer’s patients), but there is little evidence to suggest, given the current regulatory definition for research, that IRBs are weighing any direct benefits, let alone those in which there is also concomitantly little risk. Fourth, it may well be that IRBs are far too uninformed regarding newer models of research, particularly those that rely on phenomenological philosophy, or those viewed through the lenses of theoretical streams unknown or little understood by IRB members (e.g., feminist theory, race and ethnic theories, critical theories, or postcolonial perspectives). As a consequence, boards constituted with members uninformed about emerging theoretical and philosophical perspectives tend to fall back on definitions of research that ill serve cutting-edge research initiatives. From this perspective, IRBs may well operate, however inadvertently, to abridge academic and intellectual freedom at a time when it is already seemingly under attack. Qualitative researchers (of whatever theoretical perspective) clearly have “a horse in this race.” The level of “protectionism”2 surrounding research has clearly gone up from what was weak to what is now moderate and from what was moderate to what is now strong, particularly in the biomedical sciences. The new era of caution inhuman subjects research has clearly prompted more and more stringent oversight and review processes.3 Some of that oversight and review, however, may be unnecessary for many forms of research that demonstrate little, or minimal, risk to participants, and certainly not the risks associated with clinical trials and/or biomedical research. Researchers interested in collecting and analyzing the lived experiences and meaning-making activities of participants in teaching and learning contexts are not posing life-threatening possibilities. They are, however, creating the possibility of direct and indirect benefits to participants. This suggests two strategies. First, qualitative researchers must themselves become involved in IRB activities by agreeing to serve when asked and by participating in review processes that help to educate other board members who may be less than well informed about new theoretical formulations of research and inquiry. Second, where possible or necessary, individual researchers must seek to speak with IRB members, defend the research that they or their students are undertaking, and seek to educate IRBs more broadly concerning issues of level of risk and potential direct benefit. Without a significant turnaround in IRB understandings, qualitative, case study, phenomenological, and interpretive research is unlikely to receive afair hearing in institutions of higher education, whether in the oversight of research projects conducted by faculty, the dissertation research of doctoral students, or the preparation of a new generation of researchers via teaching professional skills. NOTES 1. Pritchard (in press) made this point also, although he did so in the context of teachers researching their own professional practice. The point can be made, however— and we are making it—that the criterion of generalizability in fact impedes a wide variety of other forms of research outside of the classroom. Using this criterion as a guideline for IRB approval of a study eliminates many qualitative inquiries, all constructivist and phenomenological studies, and virtually all action research, participatory action research, cooperative inquiries, and many, many others for which generalizability is neither sought, nor needed, nor desired. 2. Protectionism is, according to Moreno (2001), a “philosophical position in the ethics of human subjects research” that includes “moderate, strong and weak versions, framed in terms of how much discretion investigators should be allowed concerning the management of human subjects” (p. I-3). 3. In fact, some have suggested that the new strictures on research, on free speech in the classroom, and on encouraging open and sometimes divergent debate may pose an even greater risk: Academic freedom on campuses may suffer, or be abridged (see, for instance, American Sociological Association, 2003).

Due to the demands of external accountability as well as institutional interest in postsecondary improvement, assessment of student learning has become deeply embedded in institutional policies and practices. Especially in the past \*15 years, higher education officials have witnessed a growing body of literature on the assessment of learning and institutional policies that can support assessment. The bulk of these studies show that assessment allows institutions to be accountable and can improve educational performance (Aper 1993; Ewell 1997, 2002). A deeper understanding of and how to use assessment on our campuses comes at a good time; increased enrollments coupled with reduced public and private funding prompt institution leaders to carefully collect and use information that can inform academic planning and resource allocation. Because of their knowledge of higher education assessment, organizational theory, and research design, institutional research (IR) professionals often serve as collaborators with campus colleagues who may need assistance with survey design, statistical analysis, assessment of a discrete unit’s activity and/or broad institutional review. In my capacity as a professional in the Office of Institutional Research and Planning (IR&P) at a research extensive university, I received a call from a colleague asking if I would assist with the design and implementation of a research project for the Office of Undergraduate Research. After conferring with my director, we gladly offered assistance on the project. In this discussion, I share the major challenges and highlights of a project in which IR&P served as a willing partner on a highly successful research study. This project serves as an example work task in which an institutional researcher drew on theory and research to develop a strategy, to design a methodological plan, and then to provide data and information that was used in policy discussions among senior leaders. The IR professional’s familiarity with relevant theory, literature, and research methods was important, but this project’s success is also attributed to collaborative efforts that took place between staff members who worked in the Undergraduate Research Program office, the Center for Teaching and Learning, and faculty members who were part of an advisory committee and/ or whom worked with undergraduate research (UR) students. Faculty and administrative officials who develop curricular or co-curricular programs rely on theoretical and empirical research from Astin (1977, 1993), Pace (1982, 1987), Kuh et al. (2007), Tinto (1993), and other studies summarized in Pascarella and Terenzini (2005) for information related to student choice for admission, engagement in academic activities, and overall college student success. Research findings have strengthened or prompted the development of new activities such as first- and senior-year experience programs, UR, study abroad, service-learning, and living-learning communities in residence halls. Assessment requirements from accreditors have not lessened and may continue to require even more documentation of student learning outcomes. It is critical that institution officials utilize staff resources available to accomplish assessment goals, and IR officials can assist well with this need. Purpose of this Discussion and the Study on Undergraduate Research The purpose of this discussion is to describe an IR professional’s role in a collaborative research project that occurred between faculty and professional staff. The project involved the design, implementation, and reporting the findings of a sizable research study that examined how UR benefitted participating students and faculty members. Along with other colleagues, the IR&P office contributed staff knowledge and time to design, implement, and generate reports on this co-curricular activity that was important to senior administrative officials. This discussion will address: (1) the process that ensued to design and carry out the study and report on findings; (2) research methods used; (3) a brief summary of findings from the research study; and (4) the value of bringing together colleagues from multiple offices on campus to support the goals and objectives of the an activity that is deemed important by senior administrative officials. The discussion of the empirical study on UR is also important to the scholarly field of assessment because it contributes to the dearth of mixed-methods studies that document student learning outcomes, and thus contributes to the scholarship of assessment (Banta et al. 2002). Although I, as the IR professional, was the point person for development and implementation of the research study, discussions with the director of the Undergraduate Research Program and Undergraduate Research Advisory Committee were a valued and needed component that helped ensure project success. These colleagues helped me better understand the nuances of UR while I explained the pros and cons of specific research methods (e.g., cross-sectional versus longitudinal designs and the measure of student perceptions versus an objective measure of critical thinking). Through this study, we sought to better understand the ways in which students and faculty benefit from participation in UR. We examined student perceptions of skills and abilities gained, perceptions of satisfaction, and how they felt about the UR experience years later as alumni. Further, we investigated faculty perceptions of how the experience affected undergraduate students, graduate students, and themselves as academic researchers. The Growing Importance of Undergraduate Research in Baccalaureate Education During the past 15–20 years, there have been mounting calls for reform in undergraduate education. Supporting high levels of student learning have become deeply embedded in institutional practices and substantial evidence now confirms that student engagement and involvement in educational activities increase the likelihood for skill acquisition, satisfaction, and success in college (for example; Hu et al. 2008; Kuh et al. 2005). UR has gained prominence as a co-curricular program that facilitates student engagement and learning, and over the past two decades has become a standard feature of the baccalaureate experience. Many universities have increased their UR efforts in recent years, transforming the practice from ‘‘a cottage industry to a movement’’ (Blanton 2008). Recent works by Taraban and Blanton (2008), Lopatto (2009) and Hu et al. (2008) summarize many key findings that declare UR an important component of undergraduate education. Organizations such as ABET (ABET.org) who provide discipline-specific accreditation have also contributed to our understanding of student learning outcomes and how programs like UR can contribute to student achievement. The Importance of Theory and Relevant Literature With any empirical study, the researcher must frame the research questions within relevant theory that guides development of the study. The use of theory to frame a study is an important way to guide the study’s design, selection of the most relevant variables, measurement decisions, and analytic approaches to be used. Theory provides the researcher with cumulative information on a specific topic and frames a way to think about the issue broadly. Theory helps us articulate what we know about a topic, what we don’t yet know, and allows the researcher to consider new questions that might be asked. Along with theory, knowledge of relevant previous literature is also important. Wisely, Smart (2005) asserts that a central feature of exemplary research is its ability to thoroughly ground the study in appropriate research literature. A comprehensive set of citations documents the ‘‘intellectual heritage or foundation’’ of the study (Smart 2005, p. 464) as well as provides a framework and constructs that inform the study. With a firm grip on the literature, the researcher may also wish to talk with faculty colleagues with relevant expertise in the topic. Such discussions may provide additional insight or help clarify possible questions. Indeed, no one research study provides all the answers; each new endeavor in educational research provides additional evidence about the topic under investigation and contributes to the cumulative knowledge on that subject. In the study of UR described herein, for example, variables were considered that helped answer the overarching question on the benefits of UR, thus indicators of academic success such as grade point average (GPA) and student reports of how much time they spent on UR activities were included. Similarly, since an important goal of the study was to examine changes in the way students think and perceive the process of scientific inquiry, several possible variables were considered that provide a measure of cognition and a way to examine possible change in cognition over time. Knowledge of these variables and constructs comes from an understanding of the literature, and can be deepened by discussion with knowledgeable faculty colleagues who are authorities in the topic under study. We know that good research in IR rests on solid knowledge of the theory and relevant literature, yet in the crush of the busy work day, it is a challenge to find time to read the latest scholarly literature and consider its application to one’s particular context. When neglecting theory and relevant literature, IR studies may be designed without the benefit of recent knowledge of contemporary students and faculty, institutional leadership, and/or external constituent issues (such as the impact of lowered state appropriations on college and university priorities and subsequent actions). Conceptual Framework that Guided this Project The research questions and measures selected for this project were chosen to test our theoretical understanding of student participation and success in certain academic activities. The benefits of student participation in UR can be viewed through Pace’s theory on Quality of Effort (Pace 1982, 1987), and Astin’s Theory of Involvement (Astin 1977, 1993). Pace’s notions on quality of effort parallels many ideas proposed by Astin, and specifically emphasize the importance of depth of the experience as key to its benefit. Pace (1982) proposed that learning and development requires an investment of time and effort by the student. While time involves how often a student engages in an activity(s), effort consists of how fully or thoroughly the student delves into the activity. According to Ethington and Horn (2007), Pace believed the effort that students put forth is the most important determinant of academic outcomes. Like Pace’s Quality of Effort, Astin’s Theory of Involvement (1999) articulates many similar ideas, including the notion that the amount of student learning that occurs is directly proportional to the quality and quantity of student involvement. These notions of quality of effort and involvement fit well within the framework for UR, and students who put forth greater effort and time in UR will likely achieve higher outcomes. For example, UR students who work closely with faculty mentors to articulate project goals, design experiments, and test hypotheses are more likely to achieve intended benefits such as critical thinking and tolerance for ambiguity than students who do not engage in UR at all or do a superficially when serving as a lab receptionist. While the theoretical framework used for this project was critical to the study’s design, choice of methods, and variables, it is also important to acknowledge that assessment of academic programs as part of an institution’s overall effectiveness plan was also a motivator for the project and IR’s involvement. Some scholars argue that assessment efforts have increased primarily due to state requirements and accreditation mandates (Kuh and Ewell 2010), yet others remind us of the value of and need to perform assessment for institutional improvement (Astin 1991; Banta et al. 1993; Seifert 2010). A few institutions are well known for their deep and long-term commitment to assessment (including Alverno College, Keene College, Truman State University, and James Madison University), and numerous institutional reports at many institutions have documented practices and activities that benefit student learning. The recent Wabash National Study (Seifert et al. 2010b; Wabash National Study 2010) is also an example of a comprehensive, longitudinal study that examined critical factors related to student learning outcomes at 17 liberal arts institutions. All of these efforts have helped nurture the expectation that rigorous and regular methods of assessment will continue to have a prominent role in academic planning. Although institution-wide assessment will likely continue and perhaps expand, there is, however, great need to balance its return on investment (ROI). Institution-wide ‘assessment days’ or annual use of some published national instruments can be very costly. IR professionals can help senior administrators carefully consider what will be learned from some assessment projects, and how much will that one project contribute to the institution’s total assessment plan. Selection of measures and careful methodological design to ensure reliability, validity, generalizability to the full institution population (if desired) are important. For example, institution officials may wish to examine student involvement by class level, but use data from a cross-sectional survey. To make matters worse, low response rates can lead to biased results. If an institution expends high resources (funds and personnel) on one assessment project to receive limited responses from a small and unevenly distributed sample, those resources used may not produce valid results and thus would not have be good ROI. Initial Design Plans for the Research Study Based on objectives expressed by the director of URP and acknowledging my orientation toward a postpositivist framework for research studies, this project reflected the goal to identify and assess factors that influence the outcomes of participation in UR. In general, research from postpositivist perspective is based on observation and measurement of an objective reality (Creswell 2009), thus I was drawn to consider quantifiable measures of student and faculty gains from UR. Most IR professionals have the opportunity to get involved in survey research and analysis of small and large data sets. Cross-sectional designs are common in IR work, in part because they are less time and resource intense, yet cross-sectional designs do not control for exogenous variables such as the effect of socialization on the college experience (Pascarella 2006; Seifert et al. 2010a). A random assignment of students across treatment conditions (such as assignment to UR activities versus no UR involvement) is generally not done, because officials prefer to not restrict students from a potentially beneficial activity, as well as the heavy time and costs required for the longitudinal design. However, designs that do not have random assignment introduce the possibility of bias and possibly lower validity in the results. Indeed, it is a balance of choices requiring the IR researcher to choose a practical approach, fully aware of the potential limitations that may exist. Because limitations often exist in educational research design, I kept in mind that the ability to triangulate data across different facets of the study might prove helpful when explaining findings. Ideally, data triangulation provides a way for the researcher to connect and validate findings in different facets of a study. However, the use of multiple data points to examine research questions is also important because it signals careful consideration and thoroughness by the researcher which can bring credibility to the project and facilitate program or institutional decision making. In this study, survey data showed student perceptions of gains in specific academic skills, and comments made in the focus groups confirmed students’ enthusiasm for UR and how it helped them learn specific skills (such as use of equipment) and a broader and more informed way to think about scientific inquiry. These findings lent support to the continuation and even expanded offerings for UR activities on campus. After considering the research strategy for this project, I also considered the research method. Knowing that the benefits of participation in UR would likely occur across a variety of activities and might differ for certain students, I used a multifaceted, mixed methods approach. The four major components of the study were: (1) an analysis of student evaluations from a random sample of approximately 200 student evaluations (completed over a six-year period) by summer Science and Engineering Scholars to examine level of satisfaction, self-reported gains, and the perceived difficulties and drawbacks of their experience with the URP; (2) analysis of impact for faculty members through an institutionally- developed questionnaire. Data were gathered from faculty members who supervised UR students to examine their satisfaction, benefits, and challenges faced from involvement in UR; (3) a survey of alumni to gauge alumni perceptions of the benefits gained and level of satisfaction with their undergraduate experience and to specifically examine differences in perceived benefits between those alumni who were involved in UR and those who were not; and (4) a longitudinal study that examined the impact of UR for current undergraduate students. Each spring, the cohort of science and engineering students who volunteered for this study completed a battery of instruments to monitor level of critical thinking, college satisfaction, and level of educational engagement and some limited individual interviews. Mindful of good research design principles and knowledge of theory on student learning outcomes, it was clear to me that no one measure or method would adequately address the broad question on the effects of participation in UR. However, a multifaceted design would enable us to examine different aspects of the UR experience from alumni, faculty, and undergraduate student perspectives. Receipt of external funds provided the initial prompt to consider this study, and the second extramural award provided additional funds (and institutional visibility) to ensure a thorough mixed-methods design. Importance of the Research Design In general, student participants report high levels of satisfaction with their learning through the UR experience (Hakim 1998; Kremer and Bringle 1990; Mabrouk and Peters 2000; Manduca 1997) and faculty mentors of undergraduates believe that their students earn significant educational benefits from the research experience (Gates 1999; Kardash 2000). However, self-report data alone can be problematic. First, it is theoretically possible that students and faculty who know their responses are related to program evaluation may provide biased or incomplete answers in an attempt to cast the program in a favorable a light. Second, a survey completed only by program participants lacks a comparison or control group. One way to partial out possible respondent bias can be done through the inclusion of a control or comparison group. According to Campbell and Stanley (1960), internal validity is an absolute minimum in any research design and enables the researcher to determine if the experimental treatment makes a difference (in the phenomenon under study). To make a study design even stronger, Campbell and Stanley recommend the development of a design that also addresses external validity which seeks to determine if the treatment can be generalized to a broader population and setting. The research design might include a set of UR program participants matched with a set of individuals who resemble the participants except for the fact that they did not participate in the program. (It is acknowledged that matching may reduce the number of respondent characteristics that can be used for controls. Newer techniques such as propensity score matching assist in minimizing these methodologic concerns.) While human subjects procedures must be followed, participants’ knowledge of the study might bias their responses. Following this logic, all students who participated in this study were unaware of the primary purpose to assess the effect of participation in UR; the study was simply described to them as ‘a study of the undergraduate experience.’ Unaware that their responses would be used to assess the impact of UR, findings would be examined to see if the UR experience enhanced important skills and abilities to a greater extent than for those participants with no UR experience. In addition to including a comparison group, another important way to examine student skill enhancement that also strengthens the study’s design, is by studying a cohort of currently enrolled students over a period of time. Longitudinal studies can fall victim to internal or external validity flaws (Campbell and Stanley 1960), but they can be very helpful in examining individual variations in characteristics or traits, offer the opportunity to examine growth curves, and may enable the researcher to determine a causal relationship between variables (Bauer 2004; Seifert et al. 2010a). The longitudinal design with a pretest or baseline allows the researcher to identify the student’s baseline in measuring a skill or other variable (Seifert et al. 2010a, b). One other important feature of the mixed-methods design for the research study described herein was the inclusion of both subjective report (self reported perceptions) of satisfaction and skills gained as well as objective (standardized) measures of gains. Two standardized tests of critical thinking and reasoning were included to measure change in students’ mastery of skills that were previously reported by faculty and students. At the time of this study, no standard instrument had been developed to measure directly student and faculty-reported gains in abilities such as ‘‘carry out research,’’ ‘‘acquire information needed for problem solving,’’ or ‘‘analyze literature critically.’’ However, since many of the skills believed to be developed through UR are general and not tied to particular scientific or technical content knowledge, we decided to examine results from two different measures of general critical thinking and reasoning to see whether changes in scores or entering level of ability were associated with participation in UR. Research Questions and Major Components of the Empirical Study Although the primary focus of this discussion is focused on the considerations addressed in the design of a research study and the process of collaboration between IR officials and other colleagues on campus, a description of the research study and summary of findings are important to put the full discussion in a broader context. A brief discussion above addressed the expansion of UR programs as one important way to facilitate student success. This institution was a leader in that expansion, having begun its UR program in 1980. The program was staffed by a small team of a director, assistant director, administrative assistant, and student worker. The URP office received advice and guidance from a URP Advisory Group which was composed of approximately seven faculty members who were actively involved in UR. Although some regular student evaluations of the program occurred each year, there had not been a comprehensive and thorough evaluation of the UR program for some time. The IR professional’s knowledge of assessment and research design and advice from the URP Advisory Committee combined with the director’s deep knowledge of the UR to provide a strong group of invested colleagues who valued the opportunity to be involved in the study. In the late 1990s, this institution was acknowledged for its good practices in UR, and because of these achievements, received the award from the National Science Foundation called Recognition Award for the Integration of Research in Education (RAIRE). Receipt of the award enabled the director of the Undergraduate Research Program to fulfill her desire to complete a comprehensive analysis of the UR activity, and thus the prompt that initiated the call to IR&P for assistance with the UR study described herein. In the initial days of this project, the IR professional met on several occasions with the URP director, both individually and with the URP Advisory Group. The meetings helped the IR professional better understand the UR enterprise and Advisory Group member perspectives on facets of UR they believed to be important. These meetings also enabled the IR professional to share initial ideas for the research study and review potential research methods and a number of possible instruments (e.g., to measure aspects of critical thinking). Through a series of meetings, the team of officials agreed on the outline of the research plan and then asked the IR professional to complete the needed details. From those meetings, the following questions were formulated and then guided this study: 1. Over the course of their undergraduate careers, did UR participants believe that they develop greater intellectual curiosity, expend greater academic effort, gain greater scientific skills and understanding, and enhance leadership skills to a greater extent than comparable undergraduates who did not have the research experience perceive themselves to gain? 2. Over the course of undergraduate careers, did changes in critical thinking and reasoning scores differ for students who did and did not participate in UR, and were there differences by gender and major? Did personality typology affect students’ critical thinking or reasoning scores or the change in these scores from first to senior year? and 3. When reflecting back on their undergraduate experience, what skills and abilities did alumni report having received and was there a difference in the type or level of skills received for those who participated in UR? Why did faculty members participate in UR? What benefits did they perceive for themselves, graduate students, and undergraduate students from the UR experience? After Human Subjects approval was granted, we began the collection of data. Because of the scope of the project, we focused on certain components of the study immediately (e.g., beginning the longitudinal study right away), and knowing that we would address other components later. To help us keep on track, a timeline was developed and shared with other colleagues, especially the UR Advisory Committee. The Need to Incorporate Theory and Previous Research into the Design Mentioned above, it is important to design a research study that is grounded in relevant theory and literature. The current study on UR followed that plan; initial discussions and design were developed keeping in mind the theories on traditional-aged college student success and the large base of literature available on students, student engagement in academic and co-curricular activities, and actions that maximize success in college. Independent and dependent variables were chosen carefully, mindful of Smart’s (2005) caution on the need for parsimony. We know that previous studies have identified a large number of variables related to student success, yet a research plan that includes all of the possible variables may result in analytic threats such as collinearity. Such design threats would thwart success of the study and risk possible contribution to the literature. For this project, theory and literature related to critical thinking, cognitive development, social processes, and personality were especially important to the selection of measures, methods chosen, the time frame for collection of data, and interpretation of the findings. The literature on cognitive development and creativity informed us that changes in cognition and critical thinking occurs gradually and requires interactions with the environment (Csikszentmihayi 1996; Fischer 1980; Rogoff 1990), thus we knew that if we wanted to examine change in critical thinking or similar construct, we should consider measuring it over multiple semesters. Understanding these important constructs as they applied to traditional aged-college students was important to the design of the study and selection of instruments. It was my responsibility to share this information with colleagues in URP and on the advisory team and then lead a discussion on how to incorporate these constructs in a way to help address the research questions. Instruments Used The quest to identify instruments that could measure the benefits of UR participation was based on theory and research findings summarized above as well as faculty members’ belief that participation in research develops students’ behavioral skills and cognitive abilities. For instance, faculty members who took undergraduate collaborators into their research programs believed that students gained the ability to think logically about complex materials and understand scientific findings, developed intellectual curiosity and openness to new ideas, and gained self-confidence that was necessary to solve problems independently and as part of a team. Information about the participants in the longitudinal study was collected from the university’s student records database, from participants’ responses on four well-known published paper–pencil measures, and one interview protocol. A complement of instruments was used because we knew that all students in UR would not be involved in the same UR activities and may focus on differing sets of skills. In addition, theories on individuals and their development reminded us that students begin and end their UR participation with differing levels of intellectual ability and may see the world from different points of view. The complement of instruments would help us to examine student gains in UR in light of personal orientation and level of academic ability and included the College Student Experiences Questionnaire (CSEQ, 3rd edition; Pace 1984) and the Watson–Glaser Critical Thinking Appraisal (WGCTA; Watson and Glaser 1994). A broad measure of personality was included through the NEO-Five Factor Inventory (NEO-FFI; Costa and McCrae 1991), and because aspects of critical thinking were important to this study, two measures of episteme ologic thinking (King and Kitchener 1994) were also included. We used the Reasoning About Current Issues test (RCI, Wood 1997), the paper–pencil measure for reflective judgment, and the interview protocol, the Reflective Judgment Interview (RJI; King and Kitchener 1994). If desired, more detail on each instrument can be obtained from the author of this discussion or from the publisher of each instrument. A Summary of Study Results The primary goal for this discussion is to describe the IR professional’s role in the design and implementation of a collaborative research project and less about the specific findings. However, findings are important in the context of the expanded role for the IR professional. The discussion below provides a brief summary of findings from the four main components of this study of UR. Summative Content Analysis of Student Evaluations To examine the educational effectiveness of the UR experience, a content analysis was conducted on 183 randomly selected free-form evaluation letters (1–4 pages long) completed between 1985 and 1995 by UR students who participated in a 10-week summer program for engineering and science scholars. This content analysis determined 17 categories of responses that included increased technical skills, confidence, working with others, and understanding research and from where knowledge comes. Of the 183 responses, 113 students (62%) reported that they had learned more through the research experience than in standard courses. Some students felt liberated by the realism of their research experience; for instance, an electrical engineering student wrote, ‘‘My research experience allows me to be independent, flexible, and creative. It truly presents a challenge to myself that I have yet to find in any standard course.’’ In addition to those students who reported they learned more through research, another 39 students felt that their learning in research was as valuable as that in courses. Only two respondents said that they learned more through course work than through research. Almost all of the students (96%) described the advanced technical skills they had learned through their research assistance and 28% of the students commented directly on their generalized increased self-confidence. Nearly half of the responses (45%) indicated that the research experience had given the students important insight into the world of graduate study, thus aiding their own career decision and 45% of the respondents also commented on a new understanding of the importance of collaboration in research, not only with graduate students and their faculty sponsors, but also with technicians, industrial scientists and engineers, and others. The full report on the content analysis, can be found at: http://www.udel.edu/RAIRE/Content.pdf. Impact of UR for Faculty An institutionally-developed Faculty Survey was completed by 156 science and engineering faculty (a 44% response rate) to examine satisfaction, benefits, and challenges faced from involvement in UR. On average, faculty respondents who supervised URAs had worked with seven URAs and eight graduate or postdoctoral research assistants (GRAs) in the past 5 years. According to responses, there was a significant relationship between completion of the senior thesis and/or professional research presentations and perceived cognitive gains for students. Correlation analyses yielded significant relationships between these two activities (thesis and presentations) and students’ ability to solve problems independently, synthesize and use information, think logically about complex material, approach problems creatively, maintain openness to new ideas, and develop intellectual curiosity. Faculty respondents said they believed the research experience contributed substantially to cognitive and affective development including intellectual curiosity, understanding scientific findings, thinking logically about complex material, and synthesizing information from diverse sources. Nearly half of the faculty respondents (46%) who supervised URAs said they designed their research program to accommodate undergraduates, and half (50%) said that participation of research assistants was ‘important or very important in contributing to their research program. Accommodations for URAs include saving small, shorter projects, designing simpler projects, providing extra money or time and expanding exploratory research. Faculty respondents also indicated some barriers to the UR experience. Fifty-seven percent of the respondents said that the cost in term of time was ‘important’ to ‘very important’ and 41% said the financial cost was great. Thirty-nine of the respondents said they would accommodate more URAs if additional funds were available. Additional results for the faculty survey can be found in (Zydney et al. 2002a, b). Survey of Alumni An Alumni Survey was mailed to 2,444 alumni from the graduating classes of 1982 through 1997, approximately half of whom participated in UR. In a quasi-experimental design, responses allowed us to compare perceived benefits for UR without specifically asking the survey respondents about UR. Responses to the Alumni Survey were received from 996 individuals for a final response rate of 42%. Findings showed that survey respondents who had not participated in UR worked significantly more hours off campus than those with self-reported research involvement. The URP alumni reported having lived on campus longer as well as having participated to a greater extent in student government than the alumni with no research participation (overall number of respondents who participated in student government was small). URP alumni reported having spent more semesters conducting research than non-research respondents. Respondents also indicated the level of benefit they received from undergraduate activities. Involvement in UR, completion of a senior thesis, study abroad, and internship related to major were rated as most beneficial for the total group of respondents. Benefit received from participation in UR was significantly higher for alumni who participated in formal research than for those who self-reported research participation. Additional findings from the Alumni Survey can be found in Bauer and Bennett (2003). Impact of UR for Current Undergraduate Students As discussed above, knowledge of theory on cognitive development and student learning confirmed the value in examining change over time, thus at the beginning of the study, fulltime first-year students majoring in science and engineering were invited to participate in a four-year longitudinal study. Similar to the alumni survey, participants did not know this study specifically concerned UR and simply called the project a ‘study of undergraduate experiences.’ Each spring for four years, students who volunteered to participate in this study (initial N = 266) gathered to complete several instruments (described above). Phone call reminders, incentive prizes, and small cash awards ($5–20) were distributed to encourage continued participation. At the start of this longitudinal study (end of student’s first year), it was anticipated that some of the study’s participants would become involved in UR (typically at the end of the sophomore year), leaving the remaining study participants to serve as a comparison group coming from the same class and set of academic majors. By spring of the fourth year, 215 participants remained, including 191 science and engineering majors, yielding an overall subject retention rate of 81%. Approximately one half of the students participated in UR, leaving about one-half of the students to serve as a comparison group. Each spring, when students completed the instruments described above, they also reported the approximate average number of hours per week they had been involved in a variety of college activities, including UR. Student self-reports of research involvement were verified with faculty and then summed to reveal the approximate total number of hours over the four baccalaureate years each student had participated in UR. Total number of hours for participation in research over the course of the baccalaureate years ranged from 0 to 3,342. Mean number of hours for those who participated in research was 626 h (SD = 596). Students were grouped into three levels of research intensity; those with: (1) no research involvement (N = 109), (2) a moderate research involvement (1–500 h; N = 45), and (3) intensive research involvement (501 or more hours; N = 52). Students who participated in research for one summer and one academic year spent about 700 h of time in this activity. However, the median of research hours for participating students in this sample was 500 h. Table 1 shows descriptive statistics and Table 2 highlights the correlation matrix for select measures used in the longitudinal study. Overall CSEQ scores for this population were either consistent with or higher than the norm for doctorate-granting institutions. Based on normative scores presented in the respective manuals, NEO and WGCTA scores are generally consistent with those obtained from other baccalaureate students. RCI scores were slightly above norms for sophomores, consistent with norms at the junior year, and slightly below norms at the senior year. Although WGCTA scores were significantly correlated with SAT verbal scores, when repeated measures analyses were run, both with SAT total as a covariate and without, findings did not differ substantially. RCI scores were not highly correlated with SAT and thus no need to adjust as a covariate. Changes in Self-Reported Quality of Effort and Perceived Gains Responses to items on the CSEQ provided important data on perceived quality of effort and gains in academic and psychosocial skills. Repeated measures analyses of variance were used (Tabachnick and Fidell 1996) and specified a 4 (year) by 3 (levels of research) by 4 (major groups) by 2 (gender) design. All models were checked for good fit; and, except where noted below, each analysis was based on a model that fitted the data well. In accordance with factor groupings presented in the CSEQ manual, 3rd edition (1987), and individual items within each factor described in each section below, repeated measures ANOVAs were completed to examine overall differences in Quality of Effort (QE) scores and difference in gains scores for each year by three independent variables, participation in research, gender, and major/discipline. Regarding quality of effort in academic–intellectual (AI) activities, analyses showed no significant differences in reported quality of effort in AI activities by major or gender, although UR students reported greater AI effort in all 4 years than did students who did not participate in research. AI effort reported by students with intensive research experience was significantly higher than AI effort reported by those with no research experience. Thus, students who devoted more time to UR perceived themselves from the beginning to be academically focused and more self-directed intellectually than those who did no research. Change in Personality from First to Senior Year Based on literature reviewed for young adult growth and cognitive development and as described above, we included a broad measure of personality. Longitudinal study participants completed the NEO-FFI instrument each spring. Repeated measures findings revealed a significant main effect decrease (by year) in neuroticism from first to senior year (F = 15.889, p\.001, df = 2) but showed no significant difference in neuroticism by research, major, or gender. The analysis for openness to experience revealed a significant main effect increase by year from first to senior year (F = 6.391, p = .012, df = 2), yet again, no differences by research, major, or gender. Analyses for extraversion, agreeableness and conscientiousness revealed no significant changes from the first to senioryears. More detail on the longitudinal study can be found in Bauer and Bennett (2008) and Bauer and Liang (2003). Discussion Often, IR professionals are called upon to assist campus colleagues with data collection and assessment planning. This requires the integration of knowledge related to theory, research methods, data analysis, knowledge of the campus priorities, and effective communication techniques. Collaborative work with campus colleagues is important to the work of IR as a decision support function. The invitation to work on this project extended the collaborations between IR&P and the Undergraduate Research Program well beyond any previous efforts on this campus and reinforced the value of collaborative relationships with campus colleagues. In addition to working with staff members in URP, IR professionals (at this institution and others) regularly work with colleagues across campus including those in admissions, financial aid, centers for teaching and learning, honors programs, and many faculty who are involved in general education. The grant funding enabled us to design a comprehensive evaluation plan with the luxury of funds for the purchase of standardized instruments, a paid transcriptionist, and graduate student assistance for which IR staff members may not have usual access. The second grant from NSF enabled further expansion of the data collection, analyses, and distribute subsequent reports of our findings (including six peer-reviewed publications, several unpublished reports, and numerous conference presentations). Although too costly to do on an annual basis, a study like this every decade or so can be good ROI. Because no one instrument or method would adequately describe the benefits gained from participation in the Undergraduate Research Program, a multifaceted mixed-methods design was developed and implemented. Research studies that employ a longitudinal design, although becoming more popular (Seifert 2010), are relatively uncommon due to the time and resources needed. Work on the longitudinal study in this project prompted discussions with colleagues at professional conferences and resulted in a chapter in a monograph. Comprehensive mixed-methods studies like this one can also contribute to the scholarship of assessment described by Banta et al. (2002). Although far from perfect, the multifaceted mixed-methods design described herein might serve as a model for other studies that focus on UR or another aspect of student learning outcomes. The collaboration between the director and staff members in the UPR Program office, the URP Advisory Committee, and IR officials provided an important opportunity to utilize knowledge and skills that each person possessed. In particular, the IR professional had knowledge about assessment and research methods that assisted the process. The IR professional was also responsible for coordinating the analysis of all data collection and writing reports as a way to share the findings. In some instances, published papers were the result, but more frequent were smaller, unpublished reports and campus presentations that provided a balance of words and graphics to describe the findings. Because this study contained multiple facets that were completed over a five-year period, it was important to share initial finings in a timely way, as well as final reports and presentations that tied all components together. While this study could possibly have served as a dissertation, it was important in the reports that resulted from this study follow the goals on most IR studies— to not bury interested readers in too much detail (e.g., tedious tables with extensive statistical detail) yet enough to provide the rationale and goals for the study followed by clearly articulated results placed in the context of the institution’s perspective. In summarizing findings from the study itself, results showed that students who were involved in summer or year-long UR activities and alumni who participated in UR reported high satisfaction and positive academic progress as a result of scholars’ UR experience. Alumni who participated in UR also participated in other college experiences, such as study abroad, and more alumni with UR experience went on to graduate school. Science and engineering faculty members also reported personal benefits from involvement in UR, and they believed that students who participated in UR achieved benefits such as intellectual curiosity, understanding scientific findings, thinking logically about complex material, and synthesizing information from diverse sources. On several items, faculty perceptions about gains were consistent with those of the alumni, and the finding that significantly more research students progressed to graduate school indicate that UR benefits students. Findings from this study parallel many findings from other empirical research on UR by Nagda et al. (1998), Hathaway et al. (2002), Fechheimer et al. (2011), Lopatto (2004), and those summarized in Hu et al. (2008). In attempting to complement findings from the alumni and faculty surveys, the longitudinal study allowed for an examination of both self-reported and objective measures of academic growth. It provided the benefit of perspectives on growth from several vantage points, and incorporated the consideration of personality typology. Like the alumni study, the longitudinal study design minimized the possibility of respondent bias because the students did not know that the study of ‘‘academic experiences’’ in which they were participating had any direct relationship to UR. CSEQ results showed that in general, regardless of major discipline or gender, students with UR experience reported an increase in their academic efforts overall and gains in scientific and technical understanding to a greater extent than comparable students who had not undertaken research. Thus, CSEQ measures of academic effort and gains for enrolled students matched the perceptions of alumni. Findings from this study provide some support for Astin’s and Pace’s theories of involvement and quality of effort. Significantly higher gains in AI and personal-social items on the CSEQ for UR students confirm the position that the more effort one puts into an academic activity, the greater the benefit. In addition, while it could proposed that being involved in too many activities might spread one too thin, we found just the opposite. UR students in the longitudinal study and alumni who had participated in UR indicated involvement in many co-curricular activities. Consistent with Astin’s I-E-O model (1993), environmental characteristics that include supportive faculty, an institutional emphasis on active learning, and effective co-curricular activities can facilitate student success and satisfaction.

Empirical research in education typically includes a limitations and/or implications

sections. Even in smaller IR reports, limitations may be noted to ensure clarity on research method design, to ensure that findings are interpreted in light of caveats noted, and to help readers consider next steps based on the findings. It is a fine balance for the IR professional to assist in the decision support function by providing limited data and information, yet ensuring that senior officials are aware of the caveats and perhaps some suggestions for new or refined policies or programs that might address issues revealed from the study. Several limitations in the research study discussed herein are noted. First, since students and faculty in this study were from the sciences and engineering, results cannot be generalized to non-science disciplines. The faculty and alumni surveys were institutionally developed and were designed to address unique aspects of this institution’s UR Program and may not generalize to faculty and alumni in UR at other institutions. Mean SAT verbal scores of UR students were correlated with their overall GPA, and it is possible that high achieving students are drawn to UR more than lower ability students. Future studies on student benefits from UR may wish to explore motivation for participation. As with limitations, most research studies include a discussion on implications that enable to reader to consider what the current study may mean in a larger context. This project and results from the research study have implications for student learning, administrative support for future assessment plans, and for IR practitioners who might find themselves involved in a similar project in the future. Implications for Institutional Research Practitioners As this study of UR moved forward, it consumed increasing amounts of my time. The project allowed the IR&P office to grow in personnel (a part-time person to cover some of my IR&P duties and the addition of a graduate student on this project). Perhaps more importantly, this project helped affirm UR as a beneficial activity, seen as a good use of institutional resources, and enabled IR&P to grow in visibility on campus. This study was discussed in a number of campus meetings and because of the success of the study, the IR&P director or other IR staff members were asked to work on other projects with the Honors program, the Center for Teaching Effectiveness and efforts for General Education. The IR&P office was seen as a valuable resource to assist colleagues with data collection, analysis, and reporting needs. Conscious awareness of the need to remain unbiased in presentation of information is critical for the IR professional involved in any research project. As the IR professional for this specific project, I did not have a personal investment in the UR program, and I reported the results based on findings indicated, even when results did not show a positive result hoped for. Knowledge of the institution and trends in higher education enabled proactive thinking, and provision of data in the context of unique facets of the University. Involvement in such studies also requires an awareness of, and perhaps the ability to navigate, political waters that often exist on a campus. Faculty members in disciplines further away from education and social sciences may not be as familiar with methods and measures used in educational research, but nevertheless recognize rigorous research methods. Achieving credibility with colleagues across campus is important, and this will help not only the individual IR research but will help the entire IR office rise in value and credibility. Indeed, employing Terenzini’s three tiers of intelligence (1993) were critical to the success of this project. IR officials who get involved in assessment projects like this can contribute to its success by using their current, and perhaps greater, knowledge on literature related to organization theory, models of resource allocation, administrative and organizational practices in higher education, and previous assessment studies that can guide new studies. Implications for Assessment and Institutional Improvement All postsecondary assessment plans seek to improve student learning and indirectly, the institutions in which they are enrolled. Quiet often, IR offices are firmly involved in the institution’s efforts. While some assessments focus on student mastery of skills and abilities, others do not emphasize individual learning but instead examine school-or system level performance. Assessment may also include program evaluation, in which data is gathered to improve curricula and pedagogy (Ewell 2002). Tensions exist between the two primary purposes for assessment: accountability and improvement. Both are likely to remain, and both are important. Gray (2002) suggests there are advantages to defining assessment for both improvement and accountability because considered together they can allow us to objectively find answers to context-specific questions while also including many campus colleagues in discussions and decisions that ultimately build a culture of assessment. Gray also believes that defining assessment for both improvement and accountability legitimizes the use of many methods for information gathering. It is critical to get faculty members involved, both as scholars who might have expertise in quantitative and qualitative measures and collect data, but also as participants in the discussions of student learning and institutional mission. Faculty are important collaborators in discussions related to curriculum, pedagogy, goals for student intellectual and psychosocial learning, institutional mission, and how institutions might wish to respond to external constituents. Mentkowski and Loacker (2002) propose that faculty are important collaborators with assessment experts in discussions that contribute to the scholarship of assessment. Faculty members are important in discussions on what students ought to learn (in order to benefit the learner and society), what is possible for students to achieve, and what students are actually learning (linked to the teaching–learning process). Pondering these questions allow assessment findings to become new knowledge to many constituent groups—students, faculty, staff, departments, institutions, accrediting agencies, and state policy boards (Mentkowski and Loacker 2002). This research study and subsequent reports, publications, and presentations also had implications for the University’s reputation and prestige and could be seen as good ROI for both the institution and the granting agency. This project provided a way to showcase an important and highly successful co-curricular activity (UR) to faculty and staff members within the institution as well as externally to accreditors and others who may wish to model the UR program. This research study incorporated many components of a program review, and was cited in the institution’s self-study and subsequent documents for decennial accreditation. In addition, findings were shared at conferences, national meetings such as Council for UR (CUR), and to a group of scholars at NSF in Washington DC. Studies like this high-profile study of UR can provide many good opportunities for IR professionals. Lessons Learned This multi-year research project provided several important lessons learned. First, the project required, and I believe was successful, because of the many hours of discussion and consideration that were devoted to the overall study plan. An important part of the study’s success was the URP Director’s decision to include a number of knowledgeable faculty and administrative staff in all phases of the study. Collaboration with colleagues, each of whom had relevant expertise, was important for gathering knowledge as well as sharing information with a larger group of colleagues across campus. Discussions with colleagues was also incredibly important in making meaning of the results and helping to think thought how to best present the findings to ensure maximal benefits for the UR students and the program. In essence, these faculty members served as my informants, explaining the nuances of inquiry-based research, or how a faculty member might break down the steps of a lab project, carefully guiding, mentoring, and scaffolding the learning for a new UR student. Secondly, the level of careful thought and rigor that was integrated in all phases of the study design and report distribution was important. Conscious awareness and consideration of theory, previous literature, selection of the student sample, use of valid and reliable measures, appropriate statistical techniques, and public presentations of the findings were all critical to the study’s success. On numerous occasions, I was reminded of how grateful we were to have extramural funds that afforded us the opportunity to use certain instruments and to seek consultation from statistical experts and the assistance of graduate students. Being asked to present the findings from this study at the National Science Foundation headquarters and other national conferences brought valued attention to the study and confirmed a positive image of the institution. Third, even though this study was viewed positively by all members of the institution from day 1, there could have been some interest in slanting the outcome or embellishing certain findings. For the IR researcher, strict adherence to an unbiased position is critical. The IR researcher is valued for his or her three tiers of knowledge (Terenzini 1993), and must at all times be consciously aware of the need to remain objective and report information clearly, and in a way that can be used by senior level officials to make balanced and informed decisions. This is indeed a fine balance to achieve, but that line upon which the IR researcher walks to advise and not advocate is quite possible, and very much contributes to the successful research project.

Institutional theory has been an important framework within the field of sport management (Washington & Patterson, 2011) and has been invoked to understand meaningful topics such as governance (e.g., Kikulis, 2000; Leopkey & Parent, 2012), change (e.g., Riehl et al., 2019; Skille, 2011), leadership (e.g., Slack & Hinings, 1992), hegemony (e.g., McSweeney et al., 2019; Walker & Sartore-Baldwin, 2013), and globalization of sport (e.g., Carter, 2013; Silk et al., 2000). While Slack and colleagues have been credited with the initial exploration of institutions in sporting contexts (e.g., Amis et al., 2004; Kikulis, 2000; O’Brien & Slack, 2004; Silk & Amis, 2000; Slack & Hinings, 1994), Washington and Patterson (2011) argued that there is more to institutional theory than the concepts that had been traditionally been implemented within the sport management literature. Washington and Patterson (2011) review called for sport management researchers to (1) examine the creation, maintenance, and decline of institutions in sport; (2) provide a deep analysis of the sports landscape; and, (3) “better attention to both micro and macro mechanisms concerning institutional analyses” (p. 10). Their review noted that in many ways, sport management seemed to be stuck implementing classic, but perhaps also dated, institutional concepts (e.g., isomorphism) that point to institutionalization as a function of engrained structures and logics that are maintained and reproduced sans overt action or influence (Greenwood et al., 2008; Suddaby & Viale, 2011). In the years following Washington and Patterson (2011) review, sport management research has somewhat adopted modern institutional concepts. Notably, sport management scholars have begun interrogating the concept of embedded agency (Battilana, 2006) and how interested actors work to influence the institutions within which they are situated. Thus, institutional work has become a popular conceptualization in the sport management literature (e.g., Edwards & Washington, 2015; McSweeney et al., 2019; Nite, 2017). Institutional work is built upon the notion of embedded agency and was originally described as “the broad category of purposive action aimed at creating, maintaining, and disrupting institutions” (Lawrence & Suddaby, 2006, p. 216). Conceptually, the development of the institutional work perspective offered a new paradigm for institutional research as it has represented a bridge between the older “rational actor model” of institutionalism and neo-institutionalists’ affinity for embedded structure (Hampel et al., 2017; Lawrence & Suddaby, 2006). Institutional work addresses the “theoretical paradox of embedded agency” (Seo & Creed, 2002, p. 223) by outlining how people’s actions and institutions simultaneously shape and are shaped by each other (Battilana & D’Aunno, 2009). Indeed, institutional work has been useful for examining how sport institutions are constructed (e.g., Edwards & Washington, 2015; Nite, Ige et al., 2019), maintained (e.g., Nite, 2017; Riehl et al., 2019), and disrupted (e.g., Cocchiarella & Edwards, 2020; Nite & Hagan, 2017). The emergence of institutional work has been an important step in both the broader management and sport management literatures as it brought interests and agency back into focus for institutional scholars. However, we argue that institutional work burdens sport management scholars with clarifying the underlying assumptions and theoretical mechanisms of their research. For example, scholars need to define the institutions within their studies and explicate the foundational institutional constructs upon which their theorizations are built. In order to advance this theoretical lens, scholars can no longer rely on “default” assumptions of institutional theory and institutionalization (e.g., self replication and lack of overt “policing”) and must explicate important elements such as the actors, structures, logics, and interests (to name a few) within studied contexts. Further, the outcomes of institutionalization and perceptions of legitimacy should no longer be assumed nor taken-for-granted given that institutional work suggests that institutions may not be stable due to the complexity of interests and perceptions (Micelotta & Washington, 2013; Seo & Creed, 2002). Thus, we are suggesting the integration of institutional work by scholars to encourage the interrogation of long-held assumptions of institutional constructs within the context of their work. In many ways, institutional research can straddle the line between “theory” and “paradigm.” Crotty (1998) has indicated that a theoretical perspective is “the philosophical stance informing the methodology and thus providing a context for the process and grounding its logic and criteria” (p. 3). Schwandt (2001) suggested that there are four ways in which to understand the term theory within a social science context: (1) “[T]heory as a set of tested empirical generalizations” (p. 252); (2) “theory as a unified, systematic causal explanation of a diverse range of social phenomenon” (p. 252); (3) “theoretical orientations or perspectives (e.g., functionalism, symbolic, interactionism, behaviourism, phenomenology, hermeneutics, feminism, and poststructuralism) that are approaches to framing problems, solving problems, and understanding and explaining social reality” (p. 252); and, (4) “critical theory, which refers both to a way of theorizing and to the product of that theorizing” (p. 252). Conversely, paradigms have been described as means by which researchers identify patterns and explain those patterns through different types of models (Kuhn, 2012). The use of the term paradigm can essentially be understood in two ways. In the first understanding, a paradigm is a “cognitive framework – an exemplar set of shared solutions to substantive problems used by well-defined, specific community of scientist . . . both to generate and to solve puzzles in their field” (Schwandt, 2001, p. 183). While the other understanding of a paradigm is a “disciplinary matrix – commitments, beliefs, values, methods, outlooks, and so forth shared across a discipline” (Schwandt, 2001, pp. 183–184). It is important to distinguish paradigmatic from theoretical research of institutional theory in sport management as scholars seem to often blur those conceptual lines. In some ways, scholars seem to employ institutional theory paradigmatically by assuming the presence of institutionalization instead of detailing the relationships between variables or building new theoretical understandings within sport contexts. Doherty (2013) indicated that, “Using and developing good theory in sport management – that not only describes but explains – is important to the strength and continued growth of the field” (p. 10). With this paper, we examine how sport management scholars have implemented the underlying concepts of institutional work into institutional research. Whereas we address research prior to the publication of Washington and Patterson (2011) review, our primary focus was on research published after 2010 when institutional work began to emerge within the broader organization management and sport management fields. The aims of this paper are: one, to review the institutional theory research with sport management, focusing primarily on institutional work and embedded agency, and two, to provide insights into how scholars can increase both the quality and quantity of institutional research within sport management. We also outline new areas of theoretical development wherein sport contexts may contribute to greater understandings of the institutional theory that move beyond Washington and Patterson (2011) concerns of “hostile takeover.” Thus, we intend our work to be an inflection point that pushes sport management scholars toward more impactful institutional research. Foundations of institutional work To examine the utility of institutional work in sport management research, we begin by reviewing the development of the underlying theoretical foundations of the institutional work framework. It is not our intent to repeat the efforts of previous scholars who have provided extensive reviews of institutional concepts in previous articles and handbooks (e.g., Greenwood et al., 2017; Hampel et al., 2017; Lawrence & Suddaby, 2006; Washington & Patterson, 2011). Instead, we provide an abbreviated overview of the evolution of institutional theory and the development of the institutional work perspective. Institutional theory has become a dominant framework in the organizational literature (Greenwood et al., 2008). The conceptualizations and definitions of institutions have evolved since the early writing of Selznick (1957) who consider institutions as “organization[s] infused with value” (p. 17). Prior to the late 1970s, institutions were entities such as governmental or regulatory bodies and pillars of society like churches and prisons that were thought to be the purposeful creations of people trying to bring order to society (Stinchcombe, 1997). In this regard, the foundational paradigm of “old institutionalism” was rooted within the notion of rational action wherein entities are able to strategically create and manipulate institutions independent of engrained structures (DiMaggio & Powell, 1983; Selznic, 1996). Beginning in the late 1970s, the dominant view of institutions coalesced around the perspective that embedded structures, normative social expectations, and cognitive understandings beyond the boundaries of organizations subconsciously shaped collective action more so than a rational agency (DiMaggio & Powell, 1983; Friedland & Alford, 1991; Jepperson, 1991; Meyer & Rowan, 1977; Scott, 1995; Tolbert & Zucker, 1983). This “new institutionalism” or “neo-institutionalist” perspective was captured in Greenwood et al. (2008) broad definition of institutions as “more-or-less, taken-for-granted repetitive social behavior that is underpinned by normative systems and cognitive understandings that give meaning to social exchange and thus enable self-reproducing social order” (pp. 4–5). As understandings of institutions have progressed, certain key tenets have remained somewhat stable. Organizations and other entities are influenced by their social networks and the fields within which they are embedded (Hinings et al., 2017). Common understandings and structures within fields tend to become embedded and take on rule-like status, thereby becoming institutionalized (Meyer & Rowan, 1977). Conformity to institutionalized rules is generally achieved through mimetic, coercive, and normative pressures (DiMaggio & Powell, 1983). In this regard, organizations and other entities become somewhat isomorphic (Boxenbaum & Jonsson, 2017) to ultimately be perceived as legitimate (Deephouse & Suchman, 2008). Legitimacy is important for obtaining resources and other forms of support that are necessary for survival within institutional contexts (Suchman, 1995). Whereas neo-institutionalists tend to be sympathetic to the structuralist framework of institutional theory, around the turn of the 21st century, scholars started introducing the concept of “embedded agency” into the institutional literature (e.g., Battilana, 2006; Holm, 1995; Lawrence, 1999; Seo & Creed, 2002). Conversations of embedded agency revolve around understanding “how actors whose thoughts and action are constrained by institutions are nevertheless able to work to affect those institutions” (Zietsma & Lawrence, 2010, p. 189). It is important to understand that neo-institutionalism had presented institutional perpetuation as a relatively passive process of self-replication due to the power of embedded structures and engrained norms (Micelotta & Washington, 2013). The notion of self-replication seems to paint institutional actors as “cultural dopes” whose institutional fate is more or less predetermined (Hirsch & Lounsbury, 1997; Lawrence & Suddaby, 2006). The over-reliance on structuration presented numerous issues for institutionalists, the most notable being the concept of institutional change. Embedded agency and strategic action emerged, in part, to explain the processes whereby institutional change may emerge (Holm, 1995). From a structuralist standpoint, institutional change is a challenging concept because one of the key elements of institutions is that they are highly resistant to change (Greenwood et al., 2008) and “alternatives may be literally unthinkable” (Zucker, 1983, p. 5). A structural standpoint is based on the general theory of structuralism that connects elements of human culture through the relationship with the broader system. Furthermore, Tsebelis (1990) explained how structuralism can contribute to understandings of institutionalism as, “individual action is assumed to be an optimal adaptation to an institutional environment . . . therefore, the prevailing institutions (rules of the game) determine the behavior of the actors” (p. 40). However, scholars had begun to recognize that interested actors may leverage institutional characteristics, such as contradictory logics (Seo & Creed, 2002), political and social skill (Fligstein, 1997; Garud et al., 2002), and social positions (Battilana, 2006), to bring about change. The interplay of change and agency in institutional studies brought about the concept of institutional entrepreneurship that outlined how actors, primarily, influence institutional change (Battilana et al., 2009; Hardy & Maguire, 2017). Whereas the role of agency in institutionalism had largely been confined to understanding change, scholars also pointed to the role of agency within institutional maintenance (see Hardy & Maguire, 2008; Lawrence & Suddaby, 2006; Micelotta & Washington, 2013). As scholars had begun to understand institutional change, Seo and Creed (2002) theorized that institutional change may, in fact, be an eventuality. Thereby, scholars had begun to realize it was troublesome to assume the self-perpetuation of institutions (Hwang & Colyvas, 2011; Micelotta & Washington, 2013). Institutions face numerous challenges, such as competing logics (Seo & Creed, 2002), technological innovation (Nite & Washington, 2017), and external practices (Leblebici et al., 1991; Zietsma & Lawrence, 2010), that may result in institutional change. When institutional structures and arrangements are challenged, institutionalized actors tend to resist forces of change to protect and perpetuate the status quo (Micelotta & Washington, 2013; Trank & Washington, 2009). Thus, agency and praxis have been recognized as integral to the maintenance of institutions as well. Institutional work The re-introduction of agency and practice back into institutional theory were conceptualized and coined as “institutional work” (Lawrence & Suddaby, 2006). Lawrence and Suddaby (2006) defined institutional work as “the broad category of purposive action aimed at creating, maintaining, and disrupting institutions” (pg. 216). The theorization of institutional work has been an important development within institutional theory as it has accounted for both the structuralist concepts preferred by neo-institutionalists and the agency and praxis of old institutionalism (Lawrence et al., 2011). In this regard, the embedded agency became a central foundation of institutional work as scholars recognized that action is constrained by structure but structures are under constant barrage by actors seeking to legitimize their interests (Hampel et al., 2017; Zietsma & Lawrence, 2010). Lawrence and Suddaby (2006) are largely credited with coalescing agency and praxis under the umbrella of institutional work. They outlined three broad categories of institutional work as creation, maintenance, and disruption. Institutional creation is derived from institutional entrepreneurship, where the “activities of actors who have an interest in particular institutional arrangements and who leverage resources to create new institutions or to transform existing ones” (Maguire et al., 2004, p. 657). Dover and Lawrence (2010) described creation as “sets of practices that are learned, often slowly and out of necessity, and that they are truly ‘practices,’ in the sense that they belong not only to an individual but also to a community of actors” (p. 310). Lawrence and Suddaby (2006) outline nine types of creation work, broken down into three categories: political work (advocacy, defining and vesting), reconfiguring belief systems (constructing identities, changing normative associations and constructing normative networks), and altering boundaries of meaning systems (mimicry, theorizing, and educating). Institutional maintenance has been defined as work that “involves supporting, repairing, or recreating the social mechanisms that ensure compliance” (Lawrence & Suddaby, 2006, p. 230). The forms of maintenance work have been discussed with two categories – compliance to rule systems and perpetuating norms and beliefs (Lawrence & Suddaby, 2006). The first category is comprised of enabling, policing, and deterring work. The latter category involves valorizing and demonizing, mythologizing, and embedding and routinizing work. Further, the six types of institutional maintenance work primarily entail two domains: Rules and Symbols (Edwards & Stevens, 2019; Lawrence & Suddaby, 2006; Micelotta & Washington, 2013; Trank & Washington, 2009). The first category – Rules – involves “the use of forms of regulatory and legitimate authority” (Trank & Washington, 2009, p. 239), which entails the creation of regulations and standards, and the establishment and enforcement of policies to protect the institution (Edwards & Washington, 2015). This requires the maintenance of the regulatory and normative foundations that are established by institutions (Zilber, 2009). Scott (2008) refers to the regulatory processes (e.g., rule-setting, monitoring, and sanctioning) as the establishment of “rules, inspect others’ conformity to them, and, as necessary, manipulate sanctions – rewards or punishments – in an attempt to influence behavior” (pg. 52). The second category, Symbols, reinforces the normative and cognitive basis for the institution through the process on internalization (Selznick, 1957; Trank & Washington, 2009; Washington et al., 2008). The normative basis focuses on “rules that introduce perspective, evaluative, and obligatory dimension into social life” (Scott, 2008, p. 54) that is based on both values and norms (Scott, 2008). Values refer to the preferred or desirable structure or behavior of an institution, while norms specify how things should be done within the context of the values of the institution. The third form of institutional work is institutional disruption, which Lawrence and Suddaby suggested evolving from the concept of deinstitutionalization. Oliver (1992) defined deinstitutionalization as the “process by which the legitimacy of an established or institutionalized organizational practice erodes or discontinues . . . as a result of organizational challenges to or the failure of organizations to reproduce previously legitimated or taken-for-granted organizational actions” (p. 564). Similarly, institutional disruption is the work that involves “attacking or undermining the mechanisms that lead members to comply with institutions” (Lawrence & Suddaby, 2006, p. 235). Generally speaking, most disruption occurs, one, when actors feel their interests are not paramount within current institutional structures, and/or two, with the creation of new institutions (Lawrence & Suddaby, 2006). Although there is a wide array of disruption work, Lawrence and Suddaby (2006) identified three types of disruption work: disconnecting sanctions and rewards, disassociating moral foundations, and undermining assumptions and beliefs. The first type of disruption work is disconnecting sanctions and rewards. Lawrence and Suddaby (2006) defined this type as “working through state apparatus to disconnect rewards and sanctions from some set of practices, technologies, or rules” (p. 235). Most often, this type of disruption happens through the judiciary. The judiciary can disrupt an institution by invalidating the power, or undermining the assumptions of, the institution. However, this form of disruption can also be accomplished by professions that have the resources to do so (Lawrence & Suddaby, 2006). The second form of disruption work is disassociating moral foundations, which is defined as “disassociating the practice, rule or technology from its moral foundation as appropriate within a specific cultural context” (Lawrence & Suddaby, 2006, p. 235). Actors who engage in this form of work are most likely elites, who have a high level of cultural competence (Lawrence & Suddaby, 2006). The third and final form of disruption work is undermining assumptions and beliefs, which is defined by Lawrence and Suddaby (2006) as the “decreasing the perceived risks of innovation and differentiation by undermining core assumptions and beliefs” (p. 235). This form of disruption work can occur through innovative means that break assumptions established within an institution or by gradually acting in contradictory manners. The success of institutional disruption occurs when institutional actors are able to move beyond the barriers associated with maintaining the status quo and propose alternative behaviors (Lawrence & Suddaby, 2006). The actors that would be successful with this form of disruption work are those who can work in “highly original and potentially countercultural ways” (Lawrence & Suddaby, p. 238). Finally, it is important to note that scholars have offered different categorizations of institutional work. For instance, Lok and De Rond (2013) developed the concepts of custodial work, negotiation work, and reflexive normalization work from their study of Cambridge boat racing. Further, Hampel et al. (2017) suggested that institutional work entailed symbolic work, material work, and relational work. Lawrence and Suddaby’s (2006) framework was largely focused on the outcome of actions whereas Hampel et al. (2017) reconceptualized institutional work by examining the different mechanisms of institutional work. However, most research, especially within sport management, has largely relied on Lawrence and Suddaby (2006) conceptualizations for understanding how concerted action within institutions. Figure 1 depicts the relationships that exist between the institution, forms of an institution and the overall outcome of legitimacy work. An important by-product or overall outcome of institutional work is ensuring the legitimacy of the institution. Legitimacy is required for the creation, transformation, and diffusion of institutions, whereby other alternatives are seen as less appropriate, desirable, or viable (Dacin et al., 2002). Furthermore, Suchman (1995) indicated that legitimacy is understood to be “a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed systems of norms, values, beliefs and definitions” (pg.574). The central premise of legitimacy is that members, clients, and stakeholders perceive an organization’s actions to be desirable or proper (Edwards & Washington, 2015; Suchman, 1995). Meyer and Rowan (1977) argued that time spent by an organization in conforming to valued initiatives is representative of good management; in other words, such practices are legitimating activities (Bowerman, 2002). As institutions establish legitimacy, resources (whether financial or personnel) become more readily available to actors within the institution (Edwards et al., 2009; Slack & Hinings, 1994). Thus, legitimacy is consequential of the actors within the institution’s perception of the institutional actors (Suchman, 1995). As institutional work has gained traction within the mainstream management and organization literature, sport management scholars have begun to recognize the importance of understanding how actors influence sport institutions. In the following sections, Figure 1. An intersectoral design of an institutional work system. we review the institutional work research within the field of sport management. We then offer an overview of the advances of institutional work and how sport management scholars may effectively implement the perspective into future studies. Institutional work in sport management Since Washington and Patterson (2011) initial review of institutional theory in sport management, institutional work has gained popularity in the sport literature. Primarily, institutional work has been used in empirical settings that included national sport organizations (e.g., Dowling & Smith, 2016), NCAA (e.g., Cocchiarella & Edwards, 2020; Nite, 2017; Nite, Ige et al., 2019), anti-doping (e.g., Read et al., 2020), professional sports (Nite & Hagan, 2017; Nite et al., 2020), and the sport of hockey (e.g., Edwards & Stevens, 2019; Edwards & Washington, 2015). The inclusion of institutional work within sport settings has been an effective means for addressing Washington and Patterson (2011) concerns of deeper understandings of micro and macro processes of institutional creation, change, and maintenance. To date, much of the research that has used an institutional work lens has employed qualitative data collection methods, in the form of interviews (e.g., Edwards & Stevens, 2019; Edwards & Washington, 2015), policy document analyses (e.g., Dowling & Smith, 2016), historical content analyses (e.g., Nite, Ige, et al., 2019; Nite & Washington, 2017), and analyses of media accounts (e.g., Cocchiarella & Edwards, 2020; Nite, 2017; Nite & Nauright, 2020). Generally speaking, sport research has been largely confined within Lawrence and Suddaby (2006) forms of institutional work with Hampel et al.’s (2017) conceptualizations receiving minimal attention. As such, we outlined the following review using Lawrence and Suddaby (2006) forms. However, we recognize the limitations of breaking institutional work into specific categories. Institutional work research is often “messy” and is generally does not always fit neatly into specified categories. Indeed, Dowling and Smith (2016) noted that focusing solely on one form of work within a given study offers limited theoretical advancement as institutional work is better understood as responsive strategic activities (see also Lawrence, 1999). Thus, as we reviewed the different categories, some papers were discussed in multiple sections. Institutional creation work Institutional creation work has received attention in a variety of sport contexts. Institutional creation has been studied within emerging sports and sport organizations wherein regulatory and cognitive structures are lacking. Creation work has also been shown to be a mechanism of institutional maintenance (e.g., Nite, Ige, et al., 2019). First, Dowling and Smith (2016) detailed the case of Own the Podium (OTP) in the creation of the high-performance sport institution in Canada through the collection of “key policy documentation” (p. 401). They outlined how OTP created high-performance programs and initiatives such as the Top Secret Program, Innovation for Gold, and the Canadian Interuniversity Sport High Performance Program (also known as USPORT). The creation of these programs and initiatives seemed consistent with the institutional creation concept of constructing a normative network. Studying the emergence of Esports on US college campuses, Pizzo et al. (2019) outlined how US colleges and universities are being forced to create regulatory and cognitive institutional structures as means for organizing Esports competition and programs. Edwards and Washington (2015) also provided an account of institutional creation work by interviewing experts within the institutional setting. They described how College Hockey Inc. (CHI) emerged to enable the maintenance of collegiate hockey in the US by providing pathways from Canadian hockey players to be recruited by and compete in NCAA Division I men’s hockey programs. Also examining the sport of hockey, Li et al. (2020) showed how institutional creation work in women’s sport was a strategy for disrupting societal norms in China. Further, the research of Nite and colleagues has documented various contexts of institutional creation. Nite, Ige et al. (2019) outlined how the NCAA worked to create (and subsequently maintain; discussed subsequently) the dominant institution of college athletics in the US (see also Washington, 2004). They found that the NCAA worked with high-status universities and strategically undermined the efforts of the Amateur Athletic Union to dominate amateur sport competition in the US. They also detailed how some aspects of institutional maintenance were also a result of creation work. That is, they noted that expanding institutional boundaries and altering governance structures, both of which recreated institutional arrangements, served to maintain institutional dominance. Also studying the NCAA, Nite and Washington (2017) documented how the NCAA created governance structures for managing the emergence of television technologies. Further, Nite and Nauright (2020) discussed how sexual abuse was allowed to perpetuate on college campuses due to the creation and legitimation of discipline structures that obfuscated the investigations into and consequences of accusations of sexual abuse from athletic personnel. Finally, Nite et al. (2020) found that the creation of institutional infrastructures was an important aspect of sport entrepreneurship. Institutional maintenance work Institutional maintenance has been a popular topic within sport management. Specific contexts of maintenance work research have been mixed martial arts (MMA), college sport, and the Olympics. Sport management scholars have noted different actions that serve to maintain sport institutions. For instance, Dowling and Smith (2016) noted that hiring processes, managing institutional structures and the creation of new high performance programs were important in the maintenance of elite sport in Canada. Notably, we found that maintenance work differed depending on the level of analysis. Macro-level institutional maintenance seemed to be a process of boundary and practice expansion (see also Zietsma & Lawrence, 2010), whereas micro-level maintenance seemed more exclusionary in nature. For example, Nite, Ige, et al. (2019) macro-level account of the NCAA found that one maintenance mechanism for the NCAA was the expansion of institutional boundaries and the creation of flexible governance structures. However, Woolf et al.’s (2016) micro-level study of MMA gyms noted that institutional maintenance processes often entailed the creation of entry barriers to maintain the perceived integrity of MMA practices. Sport management scholars have supported mainstream findings showing that institutional maintenance work is resultant from threats to dominant sport institutions (see Micelotta & Washington, 2013). Agyemang et al. (2018) provided a micro-level account of institutional maintenance with their study of Tommie Smith and John Carlos’ protests at the 1968 Olympics. They found that actors with different interests, framed by different institutional logics, encountered threats to dominant institutional norms. In their study, athletes suspended their different interests and worked together to suppress the challenges and maintain the current institutional arrangements. The governance of US college sport has provided an important context for examining institutional maintenance work. Nite, Ige, et al. (2019) provided a historical analysis of the NCAA and detailed how it has maintained its dominance in US college athletics. They found that expanding institutional boundaries and restructuring institutional governance satiated challengers and allowed the NCAA to remain dominant. Nite (2017) detailed how the NCAA strategically framed its public discourse to address legitimate challenges to its governance. Drawing upon framing theory, Nite (2017) showed that media messages can be crafted to delegitimize challengers and re-establish institutional norms. Edwards’ research of hockey has also provided important accounts of institutional maintenance work. Edwards and Washington (2015) discussed how institutional creation, in the form of College Hockey Inc., was integral to the maintenance of Canadian youth hockey as it provided a legitimate means for Canadian youth hockey players to be eligible for participation in US collegiate hockey competitions. Riehl et al. (2019) noted through an analysis of interviews with coaches and board members on how a minor league hockey organization in Canada implemented some important changes to maintain the way minor league hockey was played. Additionally, they found that educational campaigns at the local level, along with advocacy of important local figures, and documented the successes of other hockey leagues served to develop support for change initiatives that ultimately served to maintain institutional arrangements in minor hockey. Edwards and Stevens (2019) examined the stakeholders who had been involved in Canadian women’s elite level hockey and the impacts those stakeholders had on the developmental pathways for elite-level players. Their research illustrated institutional maintenance through the enactment of strategic processes that “allow maintenance through legitimacy, status, and power” (pg. 11). Thus, Edwards and Stevens (2019) built upon the work of Zilber (2009), Nite (2017), and Micelotta and Washington (2013) to further illustrate the differences in regulatory processes and symbolic processes within the context of a sport setting. Data were collected through semi-structured interviews with key actors found within the institutional context of women’s hockey. As many of the empirical settings within the sport management literature have been in place for significant periods of time, it makes sense that institutional maintenance work has been popular among sport management researchers. Institutional disruption work Largely speaking, the research of institutional disruption work has been limited in scope within the field of sport management. Most institutional disruption research in sport management has focused on disruptive events and less so on the concerted efforts of interested actors to disrupt institutions. For example, Heinze and Lu (2017) outlined institutional disruption and institutional changes resultant from the widespread instances of concussions in American football. Others have documented institutional responses to technological innovations (e.g., Nite & Washington, 2017; Pizzo et al., 2019), scandals (Nite & Nauright, 2020), and disruptions from externalities (Nite, Ige, et al., 2019; Riehl et al., 2019; Woolf et al., 2016). However, few have provided specific accounts of actors engaging in institutional disruption work. Agyemang et al. (2018) documented how the protests of Tommie Smith and John Carlos at the 1968 Olympics were a form of institutional disruption. Nite and Hagan (2017) examined the media discourse of how institutional disruption may occur when institutional leaders alienate subordinates and enact values that are counter to those of subordinates and some fans. Beyond these limited accounts, there has been a significant gap in sport management research examining institutional disruption work, thus warranting further studies to continue the advancement of institutional theory within sport management. In sum, research of institutional work within the field of sport management has been somewhat limited and under-employed. Certainly, current research of institutional work in sport has been important for addressing Washington and Patterson (2011) call for extending institutional research in sport management beyond the traditions of isomorphism. Considering previous advances, institutional work could become increasingly impactful within the sport management literature as scholars continue to implement the various tenets within their research. In the next section, we outline areas wherein institutional work can be expanded within sport management research. Advancing institutional work in sport management As institutional theory continues to be popular within sport management, we contend that it is important for sport management scholars to invoke (or at least acknowledge) the most current advancements of institutional theory (i.e., institutional work). Indeed, it is important for sport management scholars to address embedded agency when studying sport institutions and other organizational phenomena. In many ways, sport management research utilizing institutional theory has done a decent job moving beyond the neo-institutionalist affinity for isomorphism and many have adopted the institutional work concepts into their studies. However, we argue that institutional work should be accounted for within all institutional studies as there is ample evidence from both the sport and mainstream management literature showing that institutions are not simply products of self-reproducing social structures. In this section, we point toward some potential gaps of knowledge that sport management scholars should consider as they seek to employ and advance institutional theory within their work. Notably, we recognize five key areas of inquiry: multi-level institutional work, legitimacy work, institutional logics, identity work, and theoretical development and integration. We expand on each subsequently. Multi-level institutionalism First, we point to the utility of adopting a multi-level approach to studying institutional work. Multi-level accounts are not new to sport management and have been used within domains such as sport participation (Hallmann et al., 2012; Wicker, Hallmann et al., 2013), volunteerism (Swierzy et al., 2018; Wicker & Hallmann, 2013), and diversity and inclusion (Cunningham, 2009, 2010, 2015; Melton & Cunningham, 2014). Institutional scholars have also adopted multi-level approaches. Notably, Nite, Hutchinson, et al. (2019) and Pedras et al. (2020) examined how organizations navigate and can be impacted by logics from multiple institutional levels. We recognize the utility and importance of multi-level research as sport and sport organizations are embedded within multiple institutions that likely impact the delivery and meaning of sport. Scholars could examine how actors and logics at the macro level (e.g., NGOs, league offices) impact action and logics at me so (e.g., team, organizational) and micro levels (e.g., athlete, coach). In doing so, the research could present a more holistic view of institutionalization within sport settings. A recent example of this approach was employed by Nite et al. (2020) investigation of the development of professional rugby in the United States. Their work documented how rugby entrepreneurs navigated the interests of the national governing body, individual clubs, sponsors, and players to develop an institutional infrastructure that would support a professional rugby league. It is also important to note that adopting a multi-level approach to institutional research presents conceptual challenges to researchers. Notably, it is important for scholars to be clear in their accounts as to what institutions they are actually studying and to provide cogent accounts of the types of institutions they are studying. To illustrate, Selznick (1957) described an institution as a “natural product of social needs and pressure – a responsive, adaptive organism” (p. 5). Building on this idea of an institution, Jepperson (1991) categorized institutions in two ways: (1) a physical entity (i.e. academic institution or organization); or (2) a process (i.e., marriage or college sport). Being able to define the institution becomes an important first step in using institutional theory and institutional work. Legitimacy work Next, we believe that institutional work research within sport management should provide greater evidence of embedded agency’s impact on perceptions of legitimacy and institutionalization. Being perceived as legitimate is important as those perceptions allow access to the cognitive support and resources necessary for survival (Deephouse & Suchman, 2008). Legitimacy has been both a focal point of sport management research (e.g., Nite & Hutchinson, 2018; Sam & Tore Ronglan, 2018; Sant & Mason, 2019; Stenling & Sam, 2017; Strittmatter et al., 2018) and has been included within other sport management institutional research (e.g., Edwards & Washington, 2015; Hemme & Morais, 2021; Huml et al., 2018; Li et al., 2020). Considering legitimacy is a social construction, we contend that institutional work is imperative when examining legitimation processes. From a critical standpoint, legitimacy is largely a constructed perception of dominant actors within a given setting who make judgements regarding the proper (or improper) institutional structures and manners in which entities behave. As such, it is important to understand how actors work to create schema for interpreting the propriety of institutional structures and actions within institutional settings. Sport management research would benefit from more studies that show how institutional work impacts perceptions of legitimacy. Edwards and Washington (2015) provided a pertinent example of how institutional work impacts perceptions of legitimacy. Another example was Hemme and Morais (2021) study of how rhetoric and organizational identity may impact legitimacy. In this regard, we continue our advocacy for multi-level examinations of institutional work. Nite, Hutchinson, et al. (2019) outlined the importance of multi-level studies in their review of escalation of commitment. They showed that perceptions of legitimacy may be influenced by institutional logics and structures at different institutional levels. Indeed, actions that may seem illogical at one level may indeed be perceived as legitimate at another (see also Nite & Hutchinson, 2018). Thus, we suggest that future studies should account for institutional work at multiple institutional levels when examining legitimacy in sport settings (see also Nite, Hutchinson, et al. (2019). For example, amateur sport organizations in Canada exist at three different levels: National (macro), Provincial/Territorial (meso), and community/ club level (micro). Research is needed to examine how the interrelationships between the government, the organizations, and between organizations to understand the work of those organizations to remain legitimate for government funding while also being able to maintain the governing relationships between the different levels within amateur sport. Data would need to be collected through interviews with the application of a case study approach. Further, Nite and Nauright (2020) noted the emergence of the concept of legitimacy work (see Figure 1) in their study of sexual abuse in college athletic departments. Whereas they focused on the legitimation of illegitimate organizational responses and processes, we argue that the further development of the concept of legitimacy work is important in the development within the institutional work framework as more work that is needed to discuss all aspect of institutional work and legitimacy. Considering that legitimacy is the desired end state of efforts towards institutionalization, it makes sense that institutional work coalesces around the concept of legitimacy work. Efforts to create, maintain, and disrupt or change institutions are all aimed at establishing the legitimacy of institutions. In some ways, one could argue that all institutional work is, in fact, legitimacy work. For example, the proposed research example above would be also applicable to further explore the legitimacy work in the context of institutional work as there is a reliance on the government at the different levels (national, provincial/ territorial, and municipal) to be viewed as legitimate to continue to receive financial support. We advocate for the continued development of legitimacy work within the sport management as well as the mainstream management literatures and see such development as an important extension of institutional work. Institutional logics Sport management scholars have invoked the institutional logics framework in numerous studies (e.g., Altukhov et al., 2020; Nite et al., 2013; O’Brien & Slack, 2004; Skirstad & Chelladurai, 2011; Southall et al., 2008; Washington & Ventresca, 2008). Generally speaking, the institutional logics perspective has been used to understand the challenges of managing sport settings that are shaped by competing or contradicting ideals. Theoretically, institutional logics have been defined as “the socially constructed, historical patterns of cultural symbols and material practices, assumptions, values, and beliefs by which individuals produce and reproduce material subsistence, organize time and space, and provide meaning to their daily activity” (Thornton & Ocasio, 1999, p. 804). Institutional logics are present at multiple institutional levels (i.e. community, corporation, family, market, profession, religion, and state) and provide the underlying rationales for institutional behavior (Thornton et al., 2012). The presence of multiple logics has been described as institutional complexity (Greenwood et al., 2011) and institutional pluralism (Kraatz & Block, 2017).1 Research of institutional logics within sport management has largely examined instances of institutional pluralism and multiple logics in regard to their impact on managerial processes. O’Brien and Slack (2004) research of the professionalization of English Rugby Union represented one of the initial studies that introduced institutional logics into the sport management literature. Others have used the framework to study contradictions in domains such as college athletics (Nite & Bopp, 2017; Nite et al., 2013; Southall et al., 2008; Washington & Ventresca, 2008), club sports (Gammelsæter, 2010; Skirstad & Chelladurai, 2011), national sport governing bodies (Pedras et al., 2020) and within the domain of sport for development (Dixon & Svensson, 2019; Phillips & Newland, 2014; Svensson, 2017; Svensson & Seifried, 2017). Certainly, studies of institutional logics within sport management would benefit from the inclusion of institutional work within their analyses. Hampel et al. (2017) noted, “There may be significant insight gained by integrating the concept of institutional logics more deeply into the study of institutional work. More specifically, we argue that the concepts of logics could provide a way into understanding how actors work to shape large-scale, cross-field institutions” (pg. 573). Thus, we too advocate for the integration of institutional work and institutional logics. In fact, much of the literature invoking institutional logics already offers multiple accounts of institutional work despite not specifically naming it. For instance, Nite et al. (2013) provided an account of how actors within a collegiate athletic department decouple practices and work to reproduce dominant logics. Additionally, the work of Svensson and colleagues has routinely pointed to actors engaging in institutional work as they develop hybrid organizational strategies in sport for development organizations (Dixon & Svensson, 2019; Svensson, 2017; Svensson & Seifried, 2017). Similarly, Altukhov et al. (2020) detailed how the Kontinental Hockey League integrated Western and Eastern European logics within their organization. These examples provide compelling accounts of how embedded agency shapes and is shaped by engrained logics within sport settings. Sport management would benefit from the research of institutional work that examines: 1. How actors create, maintain or disrupt the underlying logics within sport institutions; and 2. How actors interpret and seek to leverage the institutional logics within their sport settings. The first would be important as it would interrogate questions of how institutional logics come to be. For example, scholars could examine how different actors within the NCAA worked to create the amateur sport logic of collegiate athletics in the US. Further, scholars could examine how to create and maintain institutional logics that legitimize inclusivity in sport organizations. The second point would interrogate how actors navigate the established logics within their institutional settings. For instance, researchers could consider further examining how sport markets are created, maintained, or disrupted as entrepreneurs navigate the different institutional logics within a given setting. Nite et al. (2020) offered a pertinent example of this type of study when they examined how rugby entrepreneurs were tasked with balancing the logics of rugby and of the professional sport market in the US. As sport management scholars continue to study institutional logics, we are advocating for greater attention to embedded agency and how institutional actors shape organizational responses to institutional pluralism. Sport management research could provide important insight into how embedded actors actively reproduce and/or alter the logics within their settings. In doing so, it is important for scholars to continue to theorize the institutional work and show how it results in reproduction and/or change of institutional logics. Such an approach would be valuable both theoretical development and the practical application of institutional theory concepts within sport settings. Identity work Identity development has been an important topic both within the institutional literature (see Glynn, 2017 for an extensive review). Sport scholars, particularly sociologists and psychologists, have also developed a robust literature around the topics of athlete identity (e.g., Adler & Adler, 1991; Killeya-Jones, 2005; Lally, 2007). However, the sport literature has yet to provide research examining the intersections of identity and institutional theory. We recognize this to be an important domain for theoretical advancement and one that is well suited for institutional work. Certainly, identity has been theorized in multiple ways that are rooted in different theoretical assumptions. It is beyond the scope of this paper to unpack the theoretical debates of identity (see Lok, 2020 for an overview). Regardless, identity seems to be interconnected to institutionalization and is therefore influenced by institutional work. Scholars could consider examining how individuals craft organizational identities that comply with the engrained institutional logics of sport settings. For instance, college athletic scholars could show how athletic department identities are crafted and maintained to reflect their universities’ logics and values. We recognize some potential streams of research that would be valuable for theorizing institutional work and identity development within sport organizations. Particularly, scholars could consider examining how organizational members undertake actions to exploit institutional dynamics to develop professional identities. Additionally, scholars could examine how organizational member identities develop from different institutional logics. Research could also examine the institutional work associated with developing and maintaining organizational identities. Considering the communal aspect of sport, it seems that understanding how institutions affect individual and organizational identities could be an area where sport informs broader theoretical development within institutional theory. Theoretical development and integration Finally, we advocate for greater theoretical development and integration of broader institutional tenets, including institutional work. Institutional theory and previous research of institutional work have largely been confined to descriptive accounts. That is, research documents the presence of institutions and illustrates the types of institutional work that may have taken place within institutions. Yet, sport scholars have provided few studies that show the process of how institutionalization resulted from specific actions. To further develop the utility of institutional work, and institutional theory as a whole, scholars should adopt research methods and analyses that provide empirical evidence of statistical relationships between action and institutionalization. For instance, scholars could construct studies that show relationships between institutional maintenance actions, perceptions of legitimacy, and institutionalization. Researchers could develop questionnaires from previous studies of institutional maintenance work and adopt scales of legitimacy (Alexiou & Wiggins, 2019) and institutionalization (Singhapakdi & Vitell, 2007). Similarly, researchers could adopt process research methods that show how specific actions lead to institutionalization (see Langley, 1999). This type of research would directly address Washington and Patterson (2011) concerns for sport management to contribute to the theoretical development of institutional theory and not simply co-opt known constructs within sport management research. Such an approach would be imminently valuable for sport practitioners as research could then show how practitioner action directly impacts engrained operations in sport settings. Scholars have noted that institutional work is triggered by threats (Micelotta & Washington, 2013) and/or misaligned interests (Seo & Creed, 2002). Building from the points in the previous paragraph, we contend that an important step within institutional theory would be developing models that explore the empirical relationships between institutional work and specific institutional challenges. For example, Nite (2017) outlined how threats to the core ethos of an institution resulted in concerted public discourse that discredited the challengers and reified the legitimacy of the institution. Future studies could further establish the empirical linkages between threats and the specific invocations of institutional work. Building on the work of Trail and James (2016), the advancement of theory, particularly institutional work, can be used in conjunction with another theoretical lens. Trail and James (2016) noted, most research, if it is substantial enough to have meaning and be of value to the field, likely cannot be substantiated solely by one theory . . . . We should be thinking about the prospect of two or more theories serving as guides for our work. (p. 144) Furthermore, one of the fundamental issues in theory-driven research that Trail and James (2016) indicated was that one of the first problems we consider is a failure to derive research from theory. Indeed, We all (should) know that our research must have a good theoretical basis, and that theory should guide the direction of the research. However, it seems that some of us forget this foundational point when starting the research process. (p. 143) Having research that is theory driven from multiple theories enables the development or enhancement of one theory, while other theories could be used as a paradigm to inform and shape the research. For example, institutional theory has been combined with stakeholder theory to understanding how stakeholder salience (Huml et al., 2018). Others have used tenets of institutional theory to explain the escalation of commitment in sport settings as well (Hutchinson et al., 2015; Nite & Hutchinson, 2018; Nite, Hutchinson, et al. (2019). We suggest that institutional work and other tenets of institutional theory could be useful for explaining theoretical processes in other domains such as branding, governance, leadership, inclusion and diversity, sport for development, innovation and technology, and even in applied domains such as event and facility management. For example, understanding the work of institutional actors within an inclusion and diversity context can provide insight into how a movement such as Black Lives Matter in the National Basketball League or the Women’s National Basketball League can impact or disrupt the institutional norms that exist and was used to have a social impact within society. Institutional work has broader implications than of simply focusing on organizations and their responses to institutional pressures. Indeed, institutional work is adept at explaining the underlying engrained mechanisms and cognitive structures that influence decision-making and organizational processes. The multi-level applications of institutional work enable sport management researchers to apply this work in a number of different empirical settings. We suggest that other popular theoretical frameworks invoked by sport management scholars could be informed by the inclusion of institutional work concepts. These might include but are not limited to resource dependency theory (e.g., Asselstine & Edwards, 2019; Brown & Pappous, 2018; Wicker, Vos et al., 2013), governance (e.g., Claringbould & Knoppers, 2008; Shilbury et al., 2013), and professionalization (e.g., Ruoranen et al., 2016; Shilbury & Ferkins, 2011). Institutional work could explain how and why organizations remain embroiled in resource-dependent situations. Further, governance scholars could employ institutional work as means for explaining how and why governing bodies and governance structures have remained in place for extensive periods of time (e.g., Nite and colleagues work examing the NCAA). Finally, institutional work could be valuable for explaining the efforts of building new sport leagues (e.g., Nite et al., 2020). Thus, we argue that institutional work has utility beyond studies specifically examining sport institutions. Conclusion With this paper, we detailed the development of the institutional work perspective of institutional theory and its inclusion within the sport management literature. In doing so, we pointed to some key advancements within sport management and also highlighted some potential avenues for future research of institutional work within various sport domains. Specifically, we called for research of institutional work from a multi-level approach, further examinations of institutional work and legitimacy, institutional logics, identity, and emotions. We also argued for stronger theoretical development and integration of institutional theory as a whole and of institutional work within sport management. As such, we hope that our perspectives push sport management scholars to integrate institutional theory and institutional work within their research in ways that align sport management research with modern conceptions of institutional theory. We also recognize that sport can be a unique context for examining pertinent institutional issues given the prominence of emotions and interconnected identities within sport domains. It is important for sport management to remain at the forefront of engaging institutional theory. Disclosure statement No potential conflict of interest was reported by the authors.

In a two-part article for Research in Higher Education we describe a series of organizational strategies and systematic interventions focused on institutional improvement at Indiana University. These efforts have centered on issues central to the health and vitality of most colleges and universities. The interventions focus on efforts to achieve optimal levels of new student enrollment, to improve the transition to college, and to strengthen the first-year college experience in order to enhance student success at a large public flagship university. These efforts have been guided systematically by theoretical leads and empirical findings found in higher education literature and informed by ongoing institutional research. Large research universities present special challenges when they attempt to create integrated approaches to fostering student success. Because they are organizationally complex and pride themselves on high degrees of decentralization, an ethos of collaboration must be cultivated among administrative and student support units and faculty, academic administrators, and student life professionals. This is necessary in order to blur the boundaries between academic affairs, student affairs, business affairs, and other functional units. This kind of cooperation is needed to improve the undergraduate experience. For these reasons, a key element of the implementation process has been the use of relevant higher education literature drawn from organizational theory and institutional effectiveness work (see for example Angelo and Cross, 1993, and Cohen and March, 1974) to guide our intervention strategies. This rich body of work guided many of our assumptions about the “way things get done” at universities. There has been an emphasis on the formation of a wide range of formal and informal committees and steering groups. These groups focus on a range of issues including an annual campus-wide symposium focusing on student life and the widespread use of faculty and administrative committees that guide campus policy on the freshman and sophomore year experience, as well as enrollment management issues. These efforts have provided important opportunities to infuse findings from higher education research and campus-based institutional research into the deliberations of these various committees and task forces. They have also resulted in greater understanding and support for these efforts and have given more faculty and administrators a stake in the outcomes. This is the first half of a two-part article describing the organizational events that have resulted in an opportunity to link theory, research, and practice at a large, complex research university. In particular, we examine how the higher education research literature, along with ongoing institutional research, has informed enrollment management efforts. In the second half of this two-part article, we will look at the first-year experience, the sophomore experience, campus retention efforts, and discuss the implications of these institutional enhancement efforts for institutional practice. CONTEXT In fall of 1997 the Lilly Endowment gave Indiana University a multimillion dollar grant to increase the number of graduating students. Lilly’s interest was increasing the number of college graduates in the state of Indiana in order to increase the economic competitiveness of the state. Even though freshman to sophomore retention rates and 6-year graduation rates at Indiana University were already higher than would be expected based on student characteristics at the time of matriculation, senior administrators on the campus decided that the best way to achieve this goal would be to further enhance student persistence. George Kuh, Professor of Educational Leadership and Policy Studies, was named an Associate Dean of the Faculty and asked to co-direct the Lilly Initiative. For more than 2 decades his research interests have focused on the college student experience. In recent years, his scholarship has centered on how institutional factors such as campus culture, academic policies, and student life goals and objectives can enhance the quality of student outcomes. Early in the Lilly Initiative he suggested a modification in the focus of the project to enhancing the first-year student experience. This modification enabled more faculty and professional staff to identify with the goal of student success as opposed to the more instrumental goal of simply increasing student persistence. A series of interventions were proposed that focused on a diverse set of activities. These proposals ranged from focusing more on academic issues during new student recruitment and orientation, to creating small, thematic, living learning groups for freshmen, to improving instruction and student performance in gateway courses such as math courses typically taken by first-year students. Like all large research-oriented universities, Indiana University is complex and decentralized. Efforts to focus on student success during the freshman year necessarily involved faculty in departments that offer most of the general education courses taken by first-year students. The involvement of professional staff in academic advising, admissions and orientation, study skill centers, the registrar, and student life administrators was essential to the design and implementation of the Lilly Initiative. Organizationally, the Lilly Initiative established an array of ad hoc committees to help guide the project and to expand the number of faculty and professional staff who would have a stake in the success of this effort. These committees have also come to play an important role in creating communication channels between faculty and professional staff in areas related to first-year student success where previously none existed. Unrelated to the Lilly Initiative, the campus had been experiencing small declines in enrollment in recent years. These declines never exceeded more than 3–4% of an enrollment target of 6,000 new first-year students. Nevertheless, the cumulative effect of these declines was beginning to be felt across the campus— both financially and on enrollments in selected academic programs. Historically, approximately 25–30% of each class of new students came from outside Indiana. Tuition from both resident and nonresident students represents an important source of income. Concerns about the number of students led to greater introspection about recruitment and retention of students. Thus, pragmatic concerns about enrollments and revenues proved a useful springboard for broader and more academically compelling discussions of the instructional mission. This facilitated natural linkages between the Lilly Initiative and the emerging enrollment management efforts. With the impending retirement of the director of admissions, the chancellor took the opportunity to create a new unit of enrollment services. Because of his research interests in student college choice, the effects of financial aid on access and college choice, and enrollment management, Don Hossler, Professor of Educational Leadership and Policy Studies, was asked to serve as Vice-Chancellor of the newly created Office of Enrollment Services. The Office of Enrollment Services consisted of the offices of admissions, financial aid, and orientation. The final link in this serendipitous convergence of events was an already established strong Office of Institutional Research. Over the years, the Office of Institutional Research had developed the ability to shift its priorities to match the strategic interests of the campus. When research on faculty career paths and satisfaction was a high campus priority, the Office of Institutional Research played the centerpiece role in these efforts. Later the campus shifted its focus to educational assessment, and again the Office of Institutional Research took a leadership role. Once the Lilly Initiative and the Office of Enrollment Services were created, the institutional research office moved into a central role in the areas of retention, assessing the outcomes of new academic support initiatives, market research, and financial aid modeling. Deborah Olsen, the Assistant Vice- Chancellor and Director of the Office of Institutional Research, also held an appointment as an Associate Professor of Educational Leadership and Policy Studies. No single organizational change model, “strategic planning process,” or bureaucratic model accurately captures the events, working relationships, or strategies that have been emerging at Indiana University. Elements of Cohen and March’s garbage model of academic decision making (1974), planned change, and serendipity set the stage for a series of initiatives that have been set in motion. Once set in motion, however, intentional examples of organizational learning have been part of our strategic approach to institutional improvement. Drawing from principles from systems theory (Senge, 1990), student development theory (Astin, 1993; Kuh, Vesper, and Pace, 1997; Pace, 1990; Pascarella and Terenizini, 1991), and the literature on effective institutional policies and practices (Angelo and Cross, 1993; Hossler, Bean, and Associates, 1990; Kuh, Schuh, Whitt, and Associates, 1991), an extensive use of cross functional groups of faculty, academic and student affairs administrators, students, and others was made to intentionally exert more influence on organizational and student learning polices and outcomes. In the next section of this first installment we focus on the creation of an enrollment management organization at Indiana University Bloomington. We examine the steps taken at the organizational and strategic level and how these actions were heavily influenced by both published research and ongoing institutional research efforts. A discussion of the results of the interventions concludes the first half of this two-part article. In the second installment in the next article in this issue we focus on student success and the implications for these efforts for other campuses. FOCUS ON THE CONTEXT FOR NEW STUDENT RECRUITMENT In the last 15 years a robust body of literature has emerged on factors that influence the college enrollment decisions of students (Hossler, Braxton, and Coopersmith, 1989; Paulsen, 1990; McDonough, 1997; Hossler, Schmit, and Vesper, 1998). During the same time period a large body of scholarship has also been published on the role of financial aid on both matriculation and persistence decisions of college students (Heller, 1997; Jackson, 1978; Leslie and Brinkman, 1988; St. John, Paulsen, and Starkey, 1996; Somers and St. John, 1997; St. John, 1998). This research provided the foundation for a set of strategic initiatives and ongoing research activities that have informed recruitment activities. Space precludes a detailed discussion of both the research literature and the strategies implemented in the area of new student recruitment, but they are highlighted in this section. Timing and Personalization Illustrative Initiatives In the last 15 years, marketing literature has differentiated between promoting tangible and intangible products (Levitt, 1983; Litten, 1986, Hayek and Hossler, 1999). The marketing of higher education is frequently given as an example of an intangible product. Purchasing an intangible product involves more risk on the part of the consumer. Institutions of higher education can reduce this “psychic risk” by responding in a timely fashion when prospective students raise questions or request specific information. Providing concrete and tangible information in response to specific student questions also further reduces the psychic risks. recruitment communication. The Office of Admissions had previously taken 3 to 4 weeks to process new student applications. In addition, insufficient emphasis was placed on responding in a timely fashion to student requests for information on topics ranging from academic programs to scholarships. A year later, new student applications were being processed in 1 to 2 weeks, except during peak application periods. Our goal is to reduce this even further after a new student information system is installed next year. New student requests for information are handled more expeditiously. We are using e-mail whenever possible to decrease our response time and to provide more personalized responses. Previous market research revealed that the campus has always had a reputation for being more personalized than many large public universities. During the past 2 years, however, the degree of personalization has been enhanced. Previously, during visits of prospective students to the campus, events were scheduled in a large auditorium. Students and families were invited to attend on selected weekends, typically in groups of 1,000–2,000 people. The Office of Admissions now hosts one to three receptions daily in a living room setting that holds approximately 70 people. In addition, the financial aid office has created a freshman services team that offers individualized and small group financial aid advising to the families of prospective students. These advising sessions help emphasize personalized service and address affordability issues that are critical for many families. This team continues to work throughout summer orientation helping to solve financial aid problems as they arise. Not all of these initiatives were immediately welcome by professionals in the offices of admissions and financial aid. The changes required reallocations of staff time and priorities. Change is never easy, and many of the practitioners in these offices were unfamiliar with the research in this area; not everyone was oriented toward using research to guide practice. A combination of ongoing professional development activities and the more traditional approach of using the power of senior administrators to implement change were employed to bring about needed changes in the offices of admissions and financial aid. Another part of the effort to personalize the recruitment process was the involvement of each professional school and college. Previously, most of these academic units had little contact with prospective students. Now each school has created student ambassador groups who call all prospective students after they are admitted. In addition, admitted students receive a letter from the dean of the college or school in which they plan to major. Some departments also coordinate letters sent by faculty members. Such efforts are not unusual at small private colleges, but are not typical of large public, research-oriented universities. Price Sensitivity and Financial Aid Illustrative Initiatives During the past 15 years, a growing body of literature has demonstrated the impact of net price and financial aid on the enrollment decisions of students. In his recent meta analysis, Heller (1997) again confirmed that net cost does influence the likelihood that students will enroll in a college or university. Jackson (1978) also found that receiving a financial aid award, regardless of the amount, increased the likelihood that a student would enroll in a specific college. Indeed, campus-based financial aid awards can also be viewed as a form of courtship. About 15 years ago many private colleges and universities started to use campus- based financial aid or tuition discounting to achieve enrollment and total revenue goals. Since that time, the use of this strategy in the private sector has increased dramatically. In more recent years, public universities, especially those with high percentages of nonresident students enrolled, have also started to explore the use of campus-based financial aid or tuition discounting to achieve enrollment objectives. With the assistance of Steven Brooks of the Stillwater Group, a financial aid consultant, and the Office of Institutional Research, a series of studies of the influence of campus-based financial aid on enrollment decisions was undertaken. The first step was the laborious process of merging and cleaning admissions data and financial aid data from two distinct, and aging, mainframe information systems. We discovered that campus-based financial aid was not achieving its desired effects on the size or quality of the entering class. In addition, net tuition revenue goals were not being optimized. Subsequent analyses resulted in the creation of multivariate models to predict the matriculation of resident and nonresident students. The significant variables in these models were then used in logistic regression models to examine the effects of different approaches to campus-based financial aid on the enrollment decisions. Once a range of viable financial aid options were identified, senior associate deans from each college or school were convened as an advisory group to review a range of financial aid options that appear to optimize enrollment, quality, and revenue for the campus. This process of advice and consultation is a crucial step in this progression. Large research universities are highly decentralized, and the consultative process results in each school and college being invested in the expenditure of financial aid. The consultative process also creates opportunities for senior academic administrators to discuss what they want for the campus and for individual academic units. Such discussions do not always occur on an annual basis on large campuses. This has been an interesting and sometimes contentious process. Academic administrators have found themselves caught with conflicting goals of wanting to increase the academic quality of the entering class, increase the ethnic diversity of the entering class, and spend as little institutionally funded financial aid as possible so as to optimize net tuition revenue. A wide range of econometric models were produced so that academic administrators could see both the benefits and costs of awarding varying amounts of institutionally funded financial aid. This process made apparent the real costs of achieving each of the enrollment goals being articulated during our meetings. Geodemography Illustrative Initiatives Increasingly, new student recruitment efforts are based on the use of social science and geodemographic database analyses. Colleges and universities have started to use their student information databases to better understand the characteristics of students who applied, are admitted, and enrolled (Murray, 1991). They have also looked more carefully at the location of high schools, cities, and states in which enrolling students reside to target more effectively their direct mail and prospective student receptions. In addition to the analysis of existing student databases, geodemographic databases have become available from for profit companies such as Noel-Levitz, Prizm, Claritas Corporation, and the Sequitir Corporation. These databases enable institutions to further target their recruitment efforts. They can combine information on lifestyle decisions, purchasing patterns, and even the types of colleges to which prospective students apply. This enables admissions offices to further refine their recruitment efforts by focusing on high school students who are more likely to be interested in attending specific types of colleges or universities. Indeed, in some ways these large databases have been able to operationalize the illusive concept of student institution fit. By a more careful analysis of data from our student information system and purchasing the Enrollment Planning Service (EPS), a CD-based geodemographic tool sold jointly by The College Board and the Sequitir Corporation, we were able to more effectively identify regions of cities, and even high schools where we generated a large percentage of our applications. In some cases, the Office of Admissions was well aware of these target markets. In other cases, using further analyses and EPS, we were able to identify new areas of possible enrollment growth and to schedule off-campus receptions in these new locations. The same process helped us to identify geographical markets that had not proven to be a good use of staff time and/or direct mail activities. The Office of Institutional Research has also started to use multivariate analytical techniques to examine the effectiveness of various forms of contacts with prospective students. The results are increasingly being used to help determine where the Office of Admissions should invest its scarce financial and human resources. The use of geodemography and analysis of admissions contact data encountered little resistance. Most of the admissions professionals had a sense that the office had not been spending recruitment dollars wisely. These new analytical tools were seen as a way for them to improve their effectiveness. RESULTS To date, the interventions described in the first installment of this article have worked. During the first 2, full years of our efforts, the results of using a research based and data driven approach to our recruitment efforts have achieved most of our goals. The results of these efforts reinforce the value of using existing higher education research linked with ongoing institutional research to guide institutional decision making. The impact of these efforts has been very positive, though like most organizational interventions, not completely successful. The number of students attending off-campus admissions events during the past 2 years has significantly increased. In the entering class of 1998 we experienced a 7% increase in applications and an 8% increase in admitted students. The number of new first year students increased from 6,015 in 1997 to 6,735 in 1998, a nearly 12% increase. This was easily the largest entering class in the history of the campus. The mean SAT and class rank for the entering class remained unchanged. Financial aid analyses also were effective. The number of matriculating students who received financial aid exceeded projected yields. In 1999, the second year of these efforts, positive results were also realized, although not of the same magnitude as the first year. The entering first-year class still exceeded the targets of previous years by more than 500 students, but the actual number of new students declined to 6,515. This represents a decline of 3% from 1998, but an 8 to 10% increase from new student enrollments in any of the previous 5 years prior to 1998. It was also the second largest entering class in the history of the campus. Applications increased again for the entering class of 1999 by 11%. The number of admitted students rose by 5%. Nevertheless, the number of matriculating students dipped. Again, class rank and SAT indicators remained unchanged. Over the past 2 years the costs of recruiting each new student actually declined because of the large increase in students while the admissions budget had experienced only a slight increase. In 1999 constant dollars the average net tuition paid by an incoming student rose slightly between 1997 and 1999. Thus, there was no decrease in marginal revenue to achieve enrollment goals. Prior to the creation of the Office of Enrollment Services, confidence in the offices of admissions and financial aid were waning. There was little understanding of the important role that new student orientation could play in the final stages of new student recruitment or in helping incoming students to start classes better prepared to be successful college students. Indeed, some academic units had started to establish their own offices of new student recruitment. As a result of the steps taken these past 2 years and the success of the efforts, the academic deans are now committed to a collaborative effort with the Office of Enrollment Services. They continue to devote staff time and resources to help convert admitted students to matriculated students. In addition, they look to the Offices of Enrollment Services and Institutional Research for their primary sources of information and guidance in areas related to admissions, financial aid, and new student orientation. In the second installment of this article we will shift our attention to the efforts underway as part of the Lilly Initiative. We focus on interventions undertaken to enhance the first-year experience. Theoretical and empirical leads from the higher education research literature and institutional research efforts have continued to play a central role in these efforts. Orientation programs have provided a strong link between the efforts of the Office of Enrollment Services and the Lilly Initiative. These will be explored in the next article in this issue (pp. 223–235).

In New Zealand, as in most other countries, governments actively seek an improvement in the quality of student outcomes in higher education. Quality can be measured variously: for example, as value for money, fitness for purpose and as a process leading to student transformation (Harvey & Green, 1993). Student success indicators such as retention, progression, completion, satisfaction and engagement can be appraised on all three quality measures. Institutions that graduate a specified proportion of their students can be judged to provide value for money, be fit for purpose and possibly even to provide transformative experiences. However, Radloff and Coates (2011) suggested that it is not enough to establish an institution’s quality by accounting for its attrition, completion and satisfaction rates. The quality of the way students engage with learning must also be accounted for. Coates (2005) suggested that student engagement offers a credible way to gauge the quality of experiences, as students who are engaged in their learning are more likely to succeed than students who are disengaged. According to Kinzie (2010, p. 140) this is because ‘engagement is simultaneously about students’ investment in educational activities and also about the intentional structuring and facilitation of students’ involvement in enriching learning experiences’. In their quest for accountability, institutions research both hard and soft indicators of quality. Institutional research (IR) has tended to research whole institutions. Shreeve (2011) cited the Australasian Association of Institutional Research as confirming that IR generates, interprets and creates knowledge for the whole institution. However, researchers have also investigated student engagement, for example, in disaggregated organisational units. Pike et al. (2011) found that membership of learning communities influenced how students engage. This disaggregated level approach seems closer to the emerging picture of IR predicted by Shulman (2000) who suggested that institutional research should be focused on analysing the educative experiences of learners at levels other than the institutional. This article asks whether institutional research could help improve the quality of student engagement by researching at sub-institutional disaggregated levels such as in courses bridging to higher education in New Zealand. In answer it argues two interacting propositions: one, that student engagement provides useful indicators of quality in higher education; and, two, that research into student engagement is best conducted at a sub-institutional level. Literature foundations The article uses two major foci to examine the research question. One focus is student engagement, a complex construct sheltering different research orientations that in turn perceive engagement through different lenses. Pascarella and Terenzini (1991, 2005) identify two major research orientations that investigate student learning experiences. One, they suggest, contains theories that focus on personal growth generated from within individuals. Psychological factors such as cognitive development, motivation and identity formation are examples of what interests researchers following this orientation. A second is more sociological, focusing on how environmental factors such as teaching, institutional practices and non-institutional influences impact on student learning. Under the sociological umbrella, Pascarella and Terenzini (2005) synthesise a large body of research around teachers and teaching and the influence of institutional environments. The two orientations are not completely separate. Pascarella and Terenzini acknowledge overlaps between the orientations when they discuss, for example, research into the impact on learning of students’ families and other environmental factors such as parttime work. Such person-environment influences draw on both their psychological and sociological orientations. For example, in Australia James et al. (2010) found that more than half the students surveyed thought that paid work interfered with their academic performance. The project showed that students wanted study to fit into what they were already doing; they did not want to fit their lives around study. Psychological, sociological and crossover orientations have influenced the way engagement research has been approached. The other focus is IR, which is also complex with diverse functions, structures and different ways of working. Saupe (1990) suggested that IR involves activities that provide information to support institutional planning, policy formation and decision making. Terenzini (1999) identified three tiers in what amounts to an institutional intelligence service. A foundation tier provides technical and analytical intelligence: facts about admissions, enrolments, graduation, workload, staff–student ratios, for example. A second tier offers intelligence about issues affecting the institution currently and directly. These include issues such as quality, its assurance in programme delivery, retention and completion rates, teaching effectiveness and the utility of student support systems. The third tier Terenzini calls contextual intelligence. Here, IR researches the environment the institution works in: government policies, trends in demographics and economic health of its community, for example. Although IR professional associations operate in countries such as the United States, United Kingdom, Australia and New Zealand, the way IR is organised varies from being relatively undeveloped and decentralised to being fully professional and bureaucratic (Volkwein, 2010). There is no blueprint for the way institutional research works to meet the needs of different institutions. In New Zealand, IR is less a discrete bureaucracy than in some other countries but the functions listed by Terenzini are carried out. Traditionally, IR has focused on gathering quantitative information in a whole-of-institution approach. Increasingly during the twenty-first century both the quantitative and whole-institution emphases have been supplemented by qualitative data and research at sub-institutional levels (Shulman, 2000; Contreras-Mcgavin & Kezar, 2007). Student engagement and institutional research meet in Pascarella and Terenzini’s (2005) sociological orientation by connecting the second, the issues tier in IR (Terenzini, 1999), with teaching and teachers (Kuh, 2009). Much has been written about the teacher’s role in student engagement. Zepke et al. (2010) found that out of intrinsic motivation, teaching and external influences, teaching had the greatest impact on engagement. Bryson and Hand (2007) concluded that students are more likely to engage if they are supported by teachers who establish inviting learning environments, demand high standards, challenge and make themselves freely available to discuss academic progress. A number of reports link engagement to deep learning approaches. An Australian Council for Educational Research report (2008) found that while students’ attitudes to learning varied greatly, those who engaged in higher forms of learning such as analysing, synthesising and evaluating tended to be most engaged. This finding is supported by Hockings et al. (2008), who suggested that students who are asked by teachers to reflect, question, conjecture, evaluate and make connections between ideas whilst drawing on the ideas, experiences and knowledge of others are most deeply engaged. Teachers expecting high academic standards, supporting students to achieve these standards and challenging them to ‘stretch further than they think they can’ (Kuh et al., 2005, p. 178) enhance engagement. In their study of 20 leading colleges in the USA, Kuh et al. (2005) found that ‘tough’ assessment tasks enhance rather than hinder engagement as long as such challenges are associated with swift and focused feedback. Also at the confluence of the issues tier of IR and the sociologically oriented approach to the student experience is the role of institutional support in engagement. This examines what institutions can do to engage learners. An overview is provided by Kuh et al. (2005). They found that where institutions had cultures and processes that focused on student success, fore-grounded student learning in their mission, established high expectations, aimed for continuous improvement, invested money in support services, asserted the importance of diversity and difference and prepared students for learning in higher education, students could succeed. Porter (2006) found that institutional features such as selectivity, student cohort size and student–staff ratio have significant effects on engagement. Interestingly, Porter as well as Pike et al. (2006) found that spending money by itself did not improve student engagement. What mattered was institutional culture and mission. Kezar and Kinzie (2006) found that levels of engagement were higher in institutions where the mission included statements about valuing diversity, providing appropriate academic challenges, support and active and collaborative learning. Hu and Kuh (2003) assessed the connections between institutional learning climate, student gains and engagement. They found that similar students spending similar effort engaging in similar activities while attending different institutions report making different kinds and amounts of gains. This suggests that some institutions are more learning effective; that their values and practices make a difference. In the main, engagement research has used large-scale surveys seeking generalisable results. In the United States, the National Survey of Student Engagement and the Community College Survey of Student Engagement survey hundreds of institutions. Results are used to try to improve the student experience and student outcomes (Kuh, 2009). The Australasian University Survey of Student Engagement (AUSSE), covering Australian and New Zealand universities, similarly offers an institution-wide approach to inform senior management, employees, students and other stakeholders about the extent students are engaged in learning (Australian Council for Educational Research, 2011). However, not all engagement research involves large-scale surveys. In the UK, The Higher Education Academy follows a different path. It takes a more ‘bottom up approach’. In addition to a literature review, it funded case studies instead of a national survey (Trowler & Trowler, 2010a). The review and case studies have generated key statements about student engagement under three headings: individual student engagement, student engagement with structure and process; student engagement with identity. Case studies were conducted across and within six single institutions. All have involved special populations such as non-traditional students and students with disabilities (Trowler & Trowler, 2010b). While not employing a large-scale engagement survey, the UK does use various institutional student satisfaction surveys that cover similar but not the same ground as engagement surveys (Williams & Cappuccini-Ansfield, 2007). Results from these surveys enable fine-grained analyses of results over time. Institutions use them to create improvements (Kane et al., 2008). The large-scale American and Australasian surveys do not just report multi- and single-institutional results. They are able to drill down into the results of disaggregated populations such as women, international students and students participating in certain programmes or fields of study (Chen et al., 2009). Several articles have appeared using national survey data to investigate special populations within and across institutions. For example, in the United States, Pike et al. (2011) found that membership of learning communities impacted positively on engagement; Sudirjo and Sharma (2009) established that student’s perceptions of learning and support environments varied between individual programmes; and Nelson Laird et al. (2011) showed that gender gaps in colleges were variable, influenced by course characteristics. In New Zealand, Anderson (2010) examined engagement in a variety of different courses in one institution. In short, there is evidence in the literature that student engagement is a suitable target for institutional research and that researching engagement at sub-institutional levels could enhance institutional intelligence and the quality of its performance. New Zealand has no national survey examining the student experience as does the UK, for example, with its National Student Survey. All eight universities have, however, used the AUSSE. Results across the universities were analysed (Radloff, 2011) including chapters about ethnicity, field of study and learning at a distance. The chapter dealing with field of study recognises that engagement experiences differ between fields but confines its analysis to students studying for a bachelor degree (Comer & Brogt, 2011). Information for students studying below degree level was not analysed. Yet recent policy initiatives have focused on bridging education, which is similar to access education in the UK (Hyland, 2003). In successive tertiary education strategies (Ministry of Education, 2002, 2006, 2010), the government emphasized the need to help people not yet ready to play active roles in the workforce to develop foundation skills so that they could operate successfully in a ‘knowledge society’. The second Strategy document suggested ‘all New Zealanders need a “foundation” of knowledge, skills and dispositions to support them to participate in the economy and society’ (Ministry of Education, 2006, p. 22). The knowledge, skills and dispositions required centred on literacy, numeracy, language development and basic vocational skills that would enhance work readiness and bridge students into degree qualifications. The second Strategy document judged that by 2006 bridging education had begun to move from ‘a relatively marginal position within the tertiary education system to being a core activity…’ (Ministry of Education, 2006, p. 22). While outcomes from these programmes are closely monitored, student engagement and learning experiences are not. Only a thin cache of research informs institutions about the success of bridging programmes (for example Zepke et al., 2008; Anderson 2010). Project description This article examines data obtained from one institution participating in research funded by the New Zealand Teaching and Learning Research Initiative. It used a mixed method, quantitative-dominant approach, surveying and interviewing students enrolled for the first time in nine post-school institutions in New Zealand. A case study design was adopted. According to Stake (2005, p. 443), a case study is less a method than ‘a choice of what is to be studied’. A case may be simple or complex but must be a bounded entity. In this study each institution was conceived of as a case. The case for this article is an institution that offers the broadest range of bridging courses. The primary objective for each case study was to gauge the perceptions of students using a questionnaire and, after analysing these data, to interview a selection of participants. The questionnaire contained three scales gauging perceptions about different facets of student engagement: motivation; teaching, institutional practices and approaches to social action; and non-institutional influences; as well as demographic information. The motivation scale contained 24 Likert-type items. The impact of teaching and institutional support was surveyed in 26 Likert-type items asking how important each item was for their engagement and how well the institution engaged them on that item. The non-institutional, external influences scale included 12 Likert-type items seeking levels of agreement with statements about family, employment, social, cultural and personal influences on engagement. The case study institution gathered engagement information from students enrolled in four bridging courses. Such data are important for two reasons: they enable administrators and educators to ascertain whether IR could enhance engagement by researching engagement in disaggregated institutional units, such as courses; and to provide research information on the engagement of students in this under-researched bridging area. This case study institution was the only one that had sufficient bridging courses to answer the research question posed for this article. The case study institution returned 196 valid questionnaires, a response rate of 19.9%. Data were analysed as to how well these courses engaged respondents. The courses were foundation studies with 74 responses, education with 55 responses, computing with 20 responses and hospitality with 47 responses. From only the 26-item teaching and institutional scale, the article identifies the 10 items that respondents judged to engage them least, that is the items with the 10 highest means on the 26 items measured by the Likert scale. Second, the article triangulates the single case study results with 352 responses obtained from 164 students in foundation studies, 114 in education and 74 in computing in the other case study institutions. This approach takes the engagement spotlight off the individual, the usual focus of engagement research and focuses it on teachers and institutions, a major business for institutional research. Focusing on the 10 items students perceived to be least engaging would enable IR to identify where remedial work such as professional development or policy review is most urgently needed. Student responses were analysed using SPSS to compute means, their standard deviations and significance using independent t-tests for equality of means and a one-way analysis of variance. This article is an exploratory study to investigate whether there are significant differences in how first-time students in different courses perceive their engagement and whether results could provide support for the assumption that IR should focus on smaller entities within the whole institution. It is exploratory because the numbers surveyed are small, the courses examined are not representative of all bridging education courses and the data cannot be generalised. Findings need to be treated with caution. Moreover, at a time when experimental research is seen as the gold standard (Mayer, 2005) student perception data should be treated with additional caution, although this kind of data is defended by Hu and Kuh (2003). They argue convincingly that student perception research is valid and valuable, providing questions gauge items within respondents’ experience. Further, exploratory studies are useful in that they reveal whether further and more extensive investigations could be worthwhile. Results To get a preliminary view of which items students thought the institution did least well in engaging them, the 26 items on teaching and institutional support were divided into three bands according to means calculated from survey responses. Items with means between 1.0 and 1.9 were considered to be engaging by students. Those with means between 2.0 and 2.9 were considered to be moderately engaging, while items with means between 3.0 and 3.9 were judged to be unengaging. The 10 items with the largest means (that is, least engaging) were identified and ranked. Each item was designated to either be associated with teaching or institutional support (Table 1). No means fell into the 3.0–3.9 band and so students thought all 26 items were either engaging or moderately engaging. Four of the items fell into the moderately engaging 2.0–2.9 band; the remaining 22 items students thought were engaging. Six of the 10 least well done items are associated with teaching: mean 2.36–1.86 (Table 1). Three (1, 3 and 7) of these items, were concerned with teachers recognising that students have lives and experiences outside the classroom. Two items (4 and 8) focused on mutual challenges in the general relationship between students and teachers. One item (10) commented on feedback, a very direct teaching-learning relationship. The four institutional items all focused on the availability of information and communication issues: mean 2.14–1.85. The independent t-test of equality of means was used to ascertain whether there were significant differences at the p < .05 level between the means for the four courses and the whole institution (Table 2). Items in bold font show where course means were significantly higher (lower satisfaction) than the institutional means. Items in italics had significantly lower (greater satisfaction) means than the institution. The means for the hospitality course were significantly lower (greater satisfaction) than the means for the responses from students across the whole institution. Indeed, means for all 10 items fell into the top band (1.0–1.9) of perceptions of engagement. Foundation studies students found they were better served than students in the whole institution on one item, ‘knowing how to find my way around’. Conversely the means on five items in the education course were significantly higher (lower satisfaction) than means for the whole institution. Four of these items were about teaching, one about institutional support. The mean of one item for computing students, ‘the availability of learner support services’, was worse than the institutional average. All in all, 6 of the 10 items had significantly higher means (less satisfaction) for two courses and all means for one course were significantly lower (greater satisfaction) than the institutional average. A one-way analysis of variance was used to establish whether there were significant differences between courses on item means (Table 3). The education course showed no significant differences with students in either the foundation studies or computing courses. Similarly, no significant differences were found for the foundation studies and computing courses. The hospitality results, however, were significantly different from all other courses. Differences on four items were evident for hospitality and foundation studies courses and hospitality and education students. Significant differences on three items were found between hospitality and computing students. The means for ‘teachers challenging me in helpful ways’ were significantly different for three pairings: hospitality and foundation studies, hospitality and education, hospitality and computing. The means for ‘teachers providing prompt feedback’ were significantly different for the hospitality and foundation studies courses and for the hospitality and education courses. The hospitality and education and hospitality and computing courses showed differences on the item ‘being encouraged to question teachers’ practice’. Given the significant differences between courses in the case study institution the same items were checked from courses with similar names in the other institutions. Table 4 pictures a comparison between bridging courses in the selected case study institution and all other institutions. Foundation studies, education and computing courses were offered in other institutions but hospitality courses were not. The first two columns compare the means of the case study institution and the other institutions for foundation studies, education and computing courses. Results in the third column show that differences between the case study institution and all other institutions were significant. As higher means show lower engagement, the degree of engagement of students in the case study institution was significantly lower on all items except ‘knowing how to find my way around’. When the foundation studies courses were compared, the case study institution’s means were also significantly higher (lower engagement) than the other institutions’. Similar results were observed for computing except that the difference for ‘teachers challenging me in helpful ways’ was not significant. There were no significant differences between the case study institution and others on education. Meaning making This article asks whether institutional research could help improve student engagement by researching at sub-institutional disaggregated levels, such as in courses. Results show that there are not only significant differences within institutions in the way bridging students perceive their engagement; there are also differences in the way engagement is perceived in similar subjects between institutions. This suggests that IR has a role in researching engagement in small courses and that analysis of, reflection on and action taken on such data could improve engagement. However, student satisfaction research in other countries shows that disaggregated findings are not new. The United States (Terenzini, 1999) and the UK (Harvey 1995), for example, have been researching student satisfaction, including in courses, for some time. While it is true that student satisfaction and engagement research have the improvement of student learning in common, the purposes of satisfaction and engagement research are somewhat different. Satisfaction research tends to gauge students’ satisfaction with their general experience of learning in a wide range of areas (Williams & Cappuccini-Ansfield, 2007). Student engagement research discussed here digs more specifically and more deeply into bridging students’ perceptions of how well teachers and the institution engage them in learning. The two research purposes complement each other but do not overlap to any great extent. The results from this exploratory study show that students in the four courses in the single institution case study showed significant differences in their perceptions of how well they were engaged by teaching and institutional support. One course, hospitality, showed significantly lower means (greater engagement) than the institutional average on all 10 items. Another, education, showed lower engagement than the institutional average for half of the items. Five further items showed different results with hospitality students judging performance to be better than the institutional average and education students perceiving less engagement. Three teaching items showed both higher means (less engagement) and lower means (more engagement) than the institutional average. Two institutional support items also showed such a conflicting result. There were also significant differences between different subjects in the single case study. This finding was supported by differences found between education, foundation studies and computing students in the other institutions. In short, there are significant differences between institutions; differences that did not reflect well on the case study institution. Support for this can be found in both the engagement and student satisfaction literatures. For example, Sudirjo and Sharma (2009) found differences in student perceptions of learning and support environments between programmes and commencing and returning students in an Australian university. Beachboard et al. (2011) investigated the effect of participation in learning communities on engagement and found that membership of such communities enhanced feelings of belonging, one characteristic of engagement. Williams and Cappuccini-Ansfield (2007) report that mean satisfaction scores for different subject areas can be obtained from the National Student Survey and MacDonald et al. (2007) show course differences in student satisfaction in different years at one university. As Shreeve (2011) observed, IR has favoured the large-scale use of statistics often generated through a university’s own procedures, primarily to inform an institution’s management, its policy decisions and external environment about the work of the university. A survey of various university sites confirms that observation. According to the Office of the Provost, Harvard University (2011) for example, the role of IR is to collect, synthesise and analyse institutional data to support university decision making and meet external accountability requirements. In outlining the tasks of IR, the Harvard document mentions a variety of student and staff surveys, literature reviews, academic development projects, international student admission and enrolment statistics, modelling and benchmarking and growth analyses, among others. None of these tasks require a whole-of-institution approach. IR can do more than collect, analyse and disseminate institution-wide data. Evidence-based decision making can usefully be based on data acquired from units within an institution. Indeed, mass student engagement survey instruments, like the National Survey of Student Engagement and Australasian University Survey of Student Engagement, can drill down into results from smaller units like programmes, courses and membership of student organisations. Shreeve (2011) suggested that such endeavours are becoming more common, particularly in the UK. She observed that IR is becoming increasingly associated with enhancement and development activities. It is also adopting a more catholic approach to gathering data, using both quantitative and qualitative methods. Using disaggregated analyses to gauge institutional quality, IR can link diverse knowledge obtained from a variety of sources to enhance teaching and institutional support from micro to macro levels in the organisation (St John, 2006). So in what ways can IR help improve the quality of teaching and institutional support for student engagement given the data presented in this article? First, IR can recognise that small institutional units like courses can throw up quite diverse results. This implies that results from the micro levels of the institution are important and worth having to assist evidence-based decision making, primarily at the course and programme levels but also for central management. Second, teachers, programme managers and academic developers can be informed about the state of quality in their particular areas of responsibility. Questions about why the students of the hospitality course are generally engaged and why in some areas the students of the education course are not, can be addressed by IR and staff in the course using mixed methods research approaches. Critical reflection by teachers and administrators could suggest some answers and courses of action. Third, surveys such as the one used in this analysis, as well as the National Survey of Student Engagement (NSSE) and AUSSE reveal areas of specific strengths and weaknesses. For example, two items (‘being encouraged to question teachers’ practice’ and ‘being given information on how systems work’), one in the teaching and one in the institutional support areas, were seen by students in the hospitality and education courses in quite different lights. Teachers, programme managers and administrators can reflect on reasons for the higher than average engagement of hospitality students and lower than average level of engagement among the education students. With very specific evidence at hand, a development programme can be instituted. Indeed data from both the AUSSE and NSSE surveys are now being used to instigate such developments (Comer & Brogt, 2011; National Survey of Student Engagement, 2012). Where to from here? The evidence to support the argument that institutional research would assist institutional quality by investigating student perceptions of engagement at a micro level of the institution is suggestive from this exploratory study. Courses display significant differences in the way they perceive the efforts of teachers and institutions to support student engagement. Not only do course results differ from institutional means, courses differ from each other. Quality cannot be gauged from institutional hard outcome indicators alone. Valuable information about quality and how to enhance it is found in data about student engagement in disaggregated units of an institution. These findings are supported by work being done with AUSSE and NSSE data. However, some cautions must be acknowledged. On the technical level, the findings in this article are based on small numbers. On a conceptual level the complexity underpinning the results must be recognised. Student engagement is not a simple construct and neither is teaching and institutional support. For example, two of the survey items with higher means (lack of engagement) are about teachers not recognising influences from outside the institution. To sheet this item solely to teachers without recognising external influences on student perceptions seems overly simple. Despite these caveats, the findings in this article are worth considering as is a follow- up study using more representative data. However, rather than waiting for such a study, institutional researchers might do well to research student engagement at micro level using the data as a basis for improving teaching and institutional support to engage students more in their learning. Acknowledgements We warmly thank the Teaching and Learning Research Initiative for funding this project and our research partners in the project for their commitment and contribution. Our thanks also to the editor for his useful suggestions.

It all started in an elevator 34 years ago: During the 1978 Association for Institutional Research (AIR) Forum in Houston, a hotel guest who was not part of the conference stepped into an elevator crowded with AIR members, including at least two past presidents. The guest looked at the organizational name tags and asked: ‘So, what’s institutional research?’ The question was followed first by silence, then by nervous laughter. A dozen or so floors later, the doors opened to let the guest out, and there still had been no serious answer to the question (Terenzini 1993, p. 1). That encounter led to a 1991 AIR conference paper that led to a 1993 Research in Higher Education article (Terenzini 1993) offering a conception of institutional research as an activity requiring three ‘‘tiers of organizational intelligence.’’ I’ve been asked to consider this morning whether the knowledge and skill sets I wrote about 20 years ago are still valid for institutional research work today and for the immediate future. Or have they, like so many other things in higher education, changed? The French have an expression—actually, they have a lot of ‘‘expressions’’—but the one that may answer these questions and is the theme and part of the title of my remarks is: ‘‘Plus c¸a change, plus c’est la meˆme chose.’’ Is it the case that ‘‘The more things change, the more they stay the same’’? Or is it, as that great American philosopher and wordsmith, Yogi Berra, has put it, ‘‘like de´ja` vu all over again’’? On the comfortable assumption that most of you have not memorized (or maybe even read) the article, I want to review briefly the contents of each tier and then discuss whether those ideas or views still have some utility for institutional researchers now and over the next decade. On the Nature of Institutional Research My thinking on IR came from the 15 years of IR experience I’d had to that point (a couple at Syracuse and most as the director of the IR office at the State University of New York at Albany) and from the wonderful mentors I’d had—people with a wide and deep knowledge of the history, traditions, and goals of colleges and universities, and of the cultures and values of the people who populate them. My conception of IR also comes in part from the late Cameron Fincher, my colleague, mentor, and friend at the University of Georgia. One of AIR’s founders, Cameron wrote extensively and wisely about IR, and he entitled one of his pieces ‘‘Institutional Research as Organizational Intelligence’’ (Fincher 1978). Although my thinking did not track Cameron’s closely, the ‘‘organizational intelligence’’ metaphor captured my imagination with its implication of various forms and sources of intelligence (including some of the informal and not-so-public kind), and as I thought more and more about ‘‘institutional research’’ and discussed the topic over weekly lunch conversations with Larry Jones, another good friend and colleague at UGA, I came to think of IR as requiring multiple forms and sources of ‘‘intelligence,’’ each required to be effective in IR, each different from the others, and each as a necessary but not sufficient part of an integrated hierarchy. I concluded that there were three forms of personal and professional competence, institutional understanding, and savvy necessary to be an effective IR professional. Tier 1: Technical/Analytical Intelligence The first tier of institutional intelligence—technical and analytical intelligence—has two forms. The first is the body of technical knowledge and information required to be an IR professional on any given campus. This foundational knowledge includes familiarity with the institution’s information and data structures, variable names and operational definitions, and the counting rules that are the basic building blocks of the institution’s major data systems (e.g., admissions, registration, personnel). It includes knowing what’s in the data warehouse, how to access it, and how to manipulate it. The second form of Tier 1 intelligence is analytical and includes familiarity with and skill in using the tools of social science research. This skill set consists of familiarity with the components and canons of good education and social science research, including research design, sampling, measurement, varieties of data gathering/collection, scale (and ‘‘indicator’’) development, and the full array of both quantitative and qualitative analytical methods. Add to those skills the variations required for program evaluation and outcomes assessment. This form of intelligence, however, also includes other analytical and modeling skills more frequently found in business than in the social sciences, such as the process modeling techniques used for enrollment forecasting and projections, student-flow modeling, and faculty and staffing workload analyses and projections. And if that’s not enough, this part of the toolkit also includes facility with a wide array of information technology applications such as word processing, spreadsheets, graphic presentational applications, and database management, as well as the fast-emerging and faster-changing communications media and mechanisms. As I summed it up 20 years ago—and this still applies today in its entirety: ‘‘Tier 1 organizational intelligence is fundamental and foundational. Without the higher-level forms [tiers] of intelligence, however, it has little utility or value. By itself, it consists of data without information, processes without purposes, analyses without problems, and answers without questions (Terenzini 1993, p. 4).If you’re a golfer, Tier-1 intelligence by itself is sort of like having a golf bag with all the finest clubs but no golf balls and no course to play on. Tier 1: Technical/Analytical Intelligence—Revisited So, is Tier 1 intelligence today la meˆme chose and de´ja` vu all over again, or have some important things changed in important ways? Initially, I thought my original conception of Tier-1 was still sound after 20 years. Consider the alternative, I thought: Institutional research professionals without the technical and analytical tools and skills needed to ply their trade? Some of you will recall the mid-1980s when the Southern Association of Colleges and Schools (SACS) began requiring institutions to assess their education outcomes. Some wags dubbed the SACS requirement ‘‘The Institutional Researcher Full-Employment Act.’’ Well, I thought, doing away with Tier 1 intelligence would be the equivalent of the ‘‘The Institutional Researcher Full-Un-employment Act.’’ The need for this tier of organizational intelligence is, I think, not only foundational but immutable, sort of like gravity. Without it, we would be back in the early and mid-20th century inventing our research designs, tools, and methods from scratch. None of us wants to go back. Then I thought harder about my original conception of ‘‘Tier 1’’ as (by itself) ‘‘data without information, processes without purposes, analyses without problems, and answers without questions. ’’That characterization carried none of the urgency I now feel about the downsides of the burgeoning world of communications and analytical technologies. [It is worth remembering (painful as it may be) that when I wrote the original article, the internet was three years old, laptops weighed about 15 lbs., multilevel modeling was mostly about designing apartment and office buildings, and cellphones were barely a step up on Dick Tracy and his fantastical two-way wrist radio. The younger folks here can Google that: It’s ‘‘D-i-c-k -T-r-a-c-y.’’] In my 1995 AIR keynote talk, and borrowing the underlying idea from Lyman Glenny’s 1975 AIR Forum Presidential Address (in Peterson 1985), I warned that: The danger in being preoccupied with technology is that institutional researchers will increasingly be seen as technicians, good at what they do, but having a limited perspective and understanding of important academic and administrative issues. If that happens, institutional researchers will become increasingly marginal to the making of those decisions. The information institutional researchers provide will always be important, but they will be less and less likely to be present in the president’s council meetings when alternative courses of action are evaluated, solutions negotiated, and decisions made (Terenzini 1995, p. 16). I fear the same outcome, perhaps more keenly now than 20 years ago. Rapidly emerging social media and new soft- and hardware advances in data collection and analysis are all really cool methods and technologies. But my fear is that they can become ‘‘Maslow’s Hammers’’ (note the plural).You’ve heard the expression: ‘‘When your only tool is a hammer, every problem looks like a nail.’’ I am not arguing that institutional researchers should not augment the analytical equipment that’s now in our toolboxes. Rather, my caution is to avoid choosing study topics, not because they are important, but because they will let us use one or another of those really cool, new tools. Lest institutional researchers become the marginalized technicians Lyman Glenny warned us about, we must let the substance and the importance of a question drive our topic selections and analyses. Similarly, it is vital to IR’s utility, credibility, and respect in the eyes of faculty members and administrators that we avoid capitalizing on what’s in a database just because it’s there and just because we now have the capacity for doing ‘‘big data’’ and ‘‘data analytics’’ (and basking in the reflected glory of that and other analytical capacities). The danger, in my view, is that whatever we might turn up is likely to underspecify the complexity of most important problems. I hasten to add, once again, that I am not advocating against the use of such procedures. I am recalling the admonition that ‘‘Any database, if tortured long enough, will confess to anything. ’’Institutional databases, designed as they usually are to serve multiple purposes (often primarily transactional, credentialing, and reporting) are almost always a mile wide and an inch deep, and the most sophisticated and powerful analytical tools imaginable cannot compensate for the selection (‘‘discovery’’?) of an unimportant question, a marginally relevant topic, or an inadequate database. The new-and-cool technologies can be insidiously seductive. Competent and effective institutional researchers must be clear about why the question they are trying to answer is important. The question ‘‘Why is the information important to know, and who will care?’’ must never be far from our consciousness. [How to answer those questions is the stuff of Tier 3 organizational intelligence.] Technological change has a second downside for IR and which, in my view, threatens IR offices’ ability to function effectively. I am referring to the widespread and significant declines in survey response rates. Surveys are one of the big hammers in the IR toolkit, but today, any response rate over 30 % is an achievement (in an earlier time, I thought I’d failed if I had a participation rate of less than 60 %). At least one reason for the decline in participation rates is, I think, that most of the people we want to survey are already being bombarded by electronic communications (and pleas to complete a survey or ‘‘customer feedback’’ form of some sort). A recent Pew study found that the average teenager sends 100(!) text messages per day (in Roberts 2011). Donald Roberts (a Stanford communications professor) found in a 2009 study that the typical U.S. youth (age 8–18) devoted 7.5 hours per day to using media such as TV, audio, computers, video games, print, movies, and cell- and smartphones (Roberts 2011).That’s like drinking water from a fire hose! So why should students, faculty members, or others give priority to the IR office’s request to complete a survey? I have some thoughts on what we might do, but I believe the drop in survey response rates is significantly compromising both the validity and generalizability of IR survey results. It is ironic, indeed, to think that, in an era of vastly expanded and rapid data-collection capabilities and analytical power, IR offices may well be producing less meaningful and reliable information for decision making than they were in earlier, simpler times. Do we really want to get data of questionable validity and generalizability really, really fast and easily? Or are we willing to slow things a bit to get more reliable, valid, and potentially less-dangerous information for decision making? Tier 2: Issues Intelligence Tier 2, or ‘‘Issues Intelligence,’’ involves most of the substantive problems on which technical and analytical intelligence is brought to bear. Like Tier 1, Tier-2 intelligence has both substantive and procedural or process dimensions. Substantive Tier 2 intelligence includes knowledge of the kinds of issues and decisions that middle- and upper-level administrators in functional units face. These knowledge forms can include such managerial tasks as strategic planning; enrollment goal setting; the allocation (and reallocation) of faculty, staff, financial, and facilities resources; tuition and fee-setting; salary determination and equity issues; course evaluations; program and institutional planning; outcomes assessment and program evaluation; institutional self-study; budget development and execution; and fund raising and alumni relations. The process or procedural dimension of Tier 2 intelligence involves understanding how colleges and universities in general function, including the functional responsibilities of the various boxes arrayed in an institution’s organization chart; lines of authority; the exercise of formal and informal power; institutional, faculty, staff, and student cultures, and how decisions are made. It includes understanding the bureaucratic and collegial features of college and universities (particularly the collegial aspects and the faculty culture), but even more important is a grasp of these institutions’ political dimensions and the formal and informal dynamics of power (see for example, Baldridge 1971; Berger and Milem 2000). This form of intelligence, I would argue, also includes a general knowledge of the history of American higher education. That particular form of knowledge is the basis for knowing the difference between a college or university and a state department of motor vehicles. Tier 2: Issues Intelligence—Revisited No one would deny that American higher education has changed enormously since the First National Institutional Research Forum, held in Chicago in 1961. The racial/ethnic and cultural diversity of our institutions has increased dramatically (even as imbalances persist). In loco parentis died, redefining the relationship between an institution and its faculty members and their students. That legal redefinition and other factors changed the ways we think about and structure teaching and learning (although centuries-old beliefs and structures endure), and operational rationality and flexibility grew alongside the enormous advances in computing and information technologies. Some things did not change, however, including most of the major substantive issues, the knowledge that comprises Tier 2’s ‘‘Issues Intelligence.’’ Consider the following topics: Evaluating institutions for accountability Assessing faculty workload Affirmative action Nontraditional study and students Assuring academic progress Changing human resource needs Measuring and increasing academic productivity Computer-based systems models Improving departmental management Allocating resources Inter-institutional comparisons and peer groups(Marvin Peterson) A pretty familiar list of higher education’s challenges, eh? These issues, however, were also the topics of the first 11 volumes of New Directions for Institutional Research, which published its first issue in 1974 (Bowen 1974). Looking ahead, however, it seems to me that ‘‘knowledge of the major issues or decision areas that face institutions and the people who manage them’’ (1993, p. 4) is necessary but no longer sufficient Tier-2 intelligence for IR professionals, whether IR directors or their staff members. Beyond knowing the organization chart and the functional responsibilities and areas of decision-making in each of the major divisions, IR professionals today and in the future must know something about the contents of the literatures in at least a couple of those substantive and functional areas. I am not suggesting that IR professionals should have an encyclopedic grasp of the research in a particular area (although I think that would be a good thing). I am suggesting, instead, that it is important for IR professionals to have something more than a cocktail party conversation-level grasp of what’s known in at least a couple of the functional areas. Knowing colleagues on campus or around the country who can provide a CliffsNotes guide to a particular topic or area is highly useful, but it is no substitute for a deeper substantive knowledge of the major pieces of literature in one or more areas of institutional functioning. This more extensive Tier-2 intelligence is needed for at least two reasons. First, faced with too many things to do and not enough time to do them, and absent any easily available and informed source of understanding and guidance, well-meaning but time-short administrators and faculty members will fall back on common sense, anecdote, hearsay, and personal beliefs. None of those information sources are administrators’ and faculty members’ (or IR professionals’) friends. Second, the vastly increased ease and speed of communications have led to what Peter Ewell once c as ‘‘policy-making by fax.’’ Peter’s metaphor may be dated, but his point is still valid. In the modern information technology world, good—and bad—ideas can easily and quickly ‘‘go viral.’’ But who will sound the alarm when a really bad idea for practice or policy looks like a really good idea? As I have reminded students, ‘‘Without the evidence, you’re just another somebody with an opinion.’’ The same applies to faculty members, administrators, and institutional researchers. IR professionals knowing something about the research literature is critical to helping administrators and faculty members avoid bad (and perhaps costly) decisions; it is essential if our institutions are to avoid jumping on a practice and policy bandwagons headed for nowhere. In the not-too-distant past, for example, a number of research universities raced to replicate the success of North Carolina’s Research Triangle without knowing or understanding that it had taken three prominent research universities, with the aid of farsighted North Carolina governors, legislators, and influential citizens, 15–20 years to lay the groundwork for the Triangle’s subsequent successes. Many of those opportunistic, wannabe efforts languished and failed; other, somewhat more durable ones still struggle along. Similarly, and given conditions in institutions’ internal and external environments, my summary of the procedural dimensions of Tier-2 ‘‘issues intelligence’’—understanding how colleges and universities function formally and informally, particularly the collegial and political aspects of their cultures (especially the faculty culture), is understated. Knowing how to navigate within those cultures is as critical today as it ever was, and that will not change. But it will become increasingly important as we move into the future. Deeper and more sophisticated understanding of, and practical skill in, working within the simultaneously collegial, bureaucratic, and political worlds of a college or university is vital. In Kenny Rogers’ ‘‘The Gambler,’’ the narrator advises a young, down-on-his-luck, fellow passenger ‘‘on a train bound for nowhere’’ that ‘‘If you’re gonna play the game, boy, ya gotta learn to play it right’’ (Schiltz 1978). That sound piece of advice applies as well to IR professionals (and administrators and faculty members) as to gamblers. To be effective, an IR professional must have some grasp of ‘‘how to play the game.’’ Such understanding in higher education is a complex mix of skills that I have struggled for some time (so far unsuccessfully) to describe. I think it is close to the advice of Sun Tzu, a highly successful, 6th century BC Chinese general, in his famous book, The Art of War (Tzu 1963). General Tzu advised commanders: ‘‘Know the enemy and know yourself; in a hundred battles you will never be in peril’’ (p. 84). That book and its advice have been widely adapted for business, politics, and other professions. Although IR work should not be a war-like activity, and one’s colleagues should never be considered ‘‘enemies,’’ the advice, as I take it, is to understand the people and contexts with which one is dealing. In higher education, we must understand the roles, responsibilities, values, ways of thinking, formal and personal self-interests, and matters likely to be considered in decision-making by the administrators, faculty members, students, staff members, trustees, legislative staffers, and others with whom IR professionals must work (or to whom they may provide information, analyses, and grounded advice). Knowing one’s ‘‘enemy’’ entails a form of what psychologists call ‘‘empathy,’’ although (for my purposes) I dislike the psychological and counseling implications of that term. The ‘‘empathy’’ I believe IR professionals need is a keen understanding of the people in college and university settings; what faculty, administrators, staff, students, and others value, what is important to them. It is the ability to anticipate how others will respond to a proposal, an idea or opportunity (or threat,) and whether the reactions will be positive, neutral, or negative. It is knowing what it will take to secure others’ support for a proposal or process. It is knowing how to appeal to the values and self-interests of others, knowing what the ‘‘deal-breakers’’ are likely to be (and for oneself as well as others), and to find the common ground and ‘‘win–win’’ situations. These talents are highly beneficial at all levels in an IR office whether for gaining access to needed data or in laying out alternative courses of action and their likely consequences for department heads, deans, vice-presidents, or a president. [I hasten to add that one must ‘‘know [oneself],’’ being clear about one’s own goals, values, and limits beyond which one is unwilling compromise. Tier 3: Contextual Intelligence Tier 3, ‘‘Contextual Intelligence,’’ may be the Queen Tier, the pinnacle of the pyramid, but it depends on the other two tiers to support it. The contextually intelligent IR professional not only commands the analytical and personal skill sets and understands the topical domains that comprise Tiers 1 and 2, but also understands how to blend those two intelligence sets in a detailed and nuanced grasp of the context and culture of a particular IR operation—the institution where IR professionals practice their craft. It is what Bob Pace referred to as ‘‘the knowledge of what a college is and where it has come from.’’ It includes an understanding of the institution’s historical and philosophical evolution, its faculty and organizational cultures, its informal as well as formal campus political structures and codes, governance, decision-making processes, and customs. It includes knowledge of how business is done in this particular college or university and who the key players are in both organizational and governance units…. contextual intelligence also entails a knowledge of the local, state, national, and international environments within which the institution must function and which both present it with opportunities and constrain what it can hope to accomplish or become. This category of intelligence…represents both content and methodologies tailored to a specific institutional setting where real people are preparing to make real decisions. It is the form of intelligence that earns [institutional research professionals] legitimacy, trust, and respect (Terenzini 1993, pp. 5–6). Tier 3: Contextual Intelligence—Revisited The original description of ‘‘contextual intelligence’’ strikes me now as still moderately accurate, but also as seriously parochial and naı¨ve. The description focuses primarily on knowing the particular institution an IR professional serves and little more. In the original, I referred to ‘‘legislators,’’ ‘‘governors,’’ and ‘‘state, national, and international environments,’’ but I mentioned all those groups in only a single sentence and only in a list of other relevant ‘‘environments.’’ No particular emphasis is given to understanding an institution’s external environments and the players and dynamics in them. Things have indeed changed, but in this case I think the French expression has it wrong: Circumstances today are decidedly not the same as they were in 1991. Were I to revise the original Tier-3 text today, I would give ‘‘contextual intelligence’’ a much broader focus and a much heavier emphasis on the importance of awareness and analysis of an institution’s state, national, and international environments. In my talk at the 1995 AIR Forum in Boston (Terenzini 1995), I expressed my belief that ‘‘the blush is off the higher education rose, probably forever. Colleges and universities have come under close public scrutiny, and I think it quite unlikely that colleges and universities… will ever regain the level of public and legislative trust, respect, and financial protection they enjoyed before 1970’’ (p. 17). I had in mind higher education’s ‘‘need to compete for public funds with other public services’’ (p. 17), such as prisons, hospitals, social services, transportation and highways, and the like, an environment in some ways like today’s. The unchallenged pedestal on which society had placed colleges and universities and their degrees had begun to crumble. Today’s ‘‘external environments,’’ however, are far more challenging, even threatening. The ‘‘close public scrutiny’’ of 1990s has become much more intense. Indeed, public and legislative scrutiny today is at times adversarial, occasionally even openly hostile. Think about the emergence of the National Association of Scholars and the rise to national prominence of David Horowitz, founder of Students for Academic Freedom, an organization intended to combat ‘‘leftist indoctrination’’ in colleges and universities. Think about Pennsylvania’s former Senator Rick Santorum’s campaign criticism of President Obama as ‘‘a snob’’ for trying to improve college-program completion rates for more students. Think about the public, political, and corporate challenges to (even in some quarters scorn for) the validity and credibility of science and the people who do it. Recall the Spellings Commission’s efforts to shape how and which learning outcomes should be assessed and to redefine the purposes and nature of collegiate accreditation (Commission on the Future of Higher Education 2006). Recall Judith Eaton’s 2008 AIR Forum keynote address in which she described the shift in higher education from self-regulation to government regulation and prescription of what constitutes academic quality. She predicted that there will be no going back, that government involvement in higher education accreditation in the future will only grow in scope and intensity (Eaton 2008). The environment and climate for colleges and universities have undergone some substantial and not always positive changes over the past 20 years, and I expect that this particular form of climate change will grow colder, not warmer. IR professionals must become more fully aware of our institution’s external worlds—national and international— and the forces shaping what is—or likely soon will be—happening on our campuses. IR professionals must incorporate knowledge of external environments in their Tier-3 intelligence if they are to assist their campuses in preparing for and dealing with those changes and challenges. As the aggressiveness of the domestic challenges to colleges and universities has increased, higher education has failed to come to grips meaningfully with legitimate questions being raised. That failure has spawned what I see as the current sense of urgency to do something to fix things, an urgency that is pushing U.S. higher education toward developing simple solutions for complex problems. The current ‘‘Completion Agenda’’ effort is a case in point. Without doubt, degree completion rates in the U.S. remain at unacceptably low levels, show sharp racial/ethnic and socioeconomic group differences, and despite substantial public and private investments, they have not changed much in the past 20 years. I do not disparage the efforts of individuals, foundations, and the government to boost those rates. The Completion Agenda has goals that are worthy, but which may also mask a degree of short-sightedness. Carol Geary Schneider (2012), AAC&U’s president, argues: Completion ought reliably to mean that students have demonstrated – cumulatively, over time – their acquisition of the knowledge and skills they will need for the complex and fast-changing challenges of work, citizenship, and contemporary life. Unfortunately, however, the completion agenda is steaming ahead without setting either goals or markers for educational quality. Internationally, we have largely ignored higher education institutions and systems around the world that are engaged in activities that will revolutionize higher education in those countries (and, eventually, in ours) and from whom we might learn some things that would improve our own educational effectiveness. As Cliff Adelman (2009) has put it, ‘‘We’ve had a good run, as the saying goes, but we are no longer at the cutting edge. U.S. higher education can no longer sail on the assumption of world dominance, oblivious to the creative energies, natural intelligence, and hard work of other nations…. We cannot live in a room of mirrors, claiming that we are so unique that nothing occurring beyond that room matters’’ (p. ix). I think most immediately of assessment and quality assurance efforts abroad, one of the most exciting of which is the Bologna Process, an effort now involving 47 mostly European countries. One of its foundational efforts has been the development of an overarching ‘‘qualifications framework’’ that countries are now adapting and applying in defining learning outcomes—what learners should know, understand, and be able to do to earn a given ‘‘qualification’’ (or degree) at each of three ‘‘cycles’’ (or levels roughly analogous to our associate, bachelor’s, and master’s degrees). The goal is a common, transnational understanding of what a degree represents for learning, regardless of the field of study. According to Adelman, the process is already being imitated in Latin America, North Africa, and Australia. Its manifestation in the U.S. is the ‘‘Degree Qualifications Profile’’ project (Adelman et al. 2011). Some of the campuses represented here are participants in that emerging project. Colleges and universities—individually and collectively—must respond and contribute more meaningfully to these and other relevant national and international discussions of equitable access to college, to degree completion and the benefits that accompany it, to reviewing and revising curricula and instructional approaches, to reducing costs, and to increasing educational effectiveness. Over the past 20 years, AIR and some of its members have increasingly become important contributors to discussions on important higher education issues here and abroad. My sense, however, is that substantial room remains for more and closer involvement, something that would benefit both our campuses and our profession. At the very least, campus-based IR professionals should be aware of these broader national and international issues and their potential impacts for their campuses. IR professionals, the people with the analytical, issues, and context intelligence, knowledge, and skills have a vitally important role to play in campus and national discussions of what constitutes ‘‘educational effectiveness’’ and how it can be achieved and documented meaningfully. Institutional researchers are the boundary spanners. You occupy the middle ground between, on the one hand, administrators and faculty members with limited awareness of the research on institutions and students, and, on the other hand, scholars who may have the theoretical and analytical skills but often lack the experience, tolerance for ambiguity, or understanding of higher education ‘‘on the ground.’’ Summary and Concluding Thoughts So, is ‘‘On the Nature of Institutional Research’’ really ‘‘Plus c¸a change, plus c’est la meˆme chose’’ (The more things change, the more they stay the same)? The academic in me wants to say: ‘‘It all depends.’’ The former IR director in me says: ‘‘Not so fast. A lot has changed, and there’ll be no going back.’’ Let me summarize what I think are the biggest changes in the worlds of higher education and institutional research and what I think some of the implications of those changes are for the field and profession: Tier-1 (Technical/Analytical) Intelligence The need for mastery of the IR analytical toolbox remains foundational. Without that, IR professionals are out of work. The tools have become increasingly sophisticated (not always a good thing), but ‘‘analysis’’ remains IR’s ‘‘core business,’’ and the analytical toolkit is essential. The dramatic advances in communications and analytical technologies, however, have two downsides. First, the glitter and dash of the new tools can be seductive, leading us to value tools over substance, perhaps losing sight of the fact that all the highpowered, sophisticated computing and analytical tools in the world cannot compensate for inconsequential questions and studies with limited real utility. We must be clear—first and foremost to ourselves and others—exactly why a study is important to do. Who will care? Will the juice be worth the squeeze? Downside No. 2 is that the volume of students’ and faculty members’ electronic media traffic threatens our ability to do our work with rigor. Survey research is a mainstay of the profession, and response rates are dropping steadily, along with the validity and generalizability of the data we gather. The bottom line is that declining response rates are compromising our ability to turn data into information that is useful and credible for decision-making. Indeed, this problem may be one of the most formidable and threatening challenges we face. IR professionals simply must find either ways to increase survey response rates or alternatives to surveys. Tier-2 (Issues) Intelligence Knowledge of the major operational domains and issues for colleges and universities is still necessary, but it is no longer sufficient for effective performance as an IR professional. A competent IR professional must have some knowledge of at least some of the research literature in at least a few of those operational areas (and the more the better). The knowledge explosion is pushing administrators and faculty leaders toward decision-making based on little or poor empirical evidence. In today’s high-speed, communication-rich environment, bad ideas can go-viral as quickly and easily as good ones, but who will know which is which? Only someone familiar with the knowledge base in a particular topic area will know the difference and—with IR professionals’ access to administrators and faculty leaders—be in a position to provide informed analysis and sound advice on alternative courses of action. Tier-3 (Contextual) Intelligence This form of organizational intelligence may be the ‘‘Queen Tier,’’ but we need to understand it more broadly than I suggested originally. Contextual intelligence must move beyond the campus boundaries, beyond a parochial knowledge of the culture, values, and traditions of our particular institution and how to function successfully in it. Understanding the local context must be augmented by at least some awareness and understanding of state, regional, national, and international contexts, issues, and events, as well as their potential implications for our campuses. Understanding ‘‘how to play the game’’ locally is still important, but it is more important now than previously to understand both what the game is beyond our campus and what’s needed for our institution to play it effectively. Finally, the widespread and sometimes hostile challenges to higher education have brought defensiveness and the conviction that an urgent response—Something. Anything!— is needed. The sense of urgency pushes decision- and policy-makers toward simple solutions for complex problems. We latch-on to easy and readily available metrics of productivity and an ill-defined ‘‘quality.’’ The urgency of responding is trumping serious consideration of effectiveness. In responding quickly, we risk trivializing the complexity of our missions and the processes needed to fulfill them. AIR and its members must be prepared to contribute to discussions about responsible changes that are needed on our campuses and nationally if higher education is to deliver on what we promise in our admissions materials. Meaningful responses to the challenges are also needed if we are to restore the public’s trust and confidence in our institutions. I believe IR professionals have important contributions to make in recovering that trust and confidence. Cameron Fincher (1985) wrote that ‘‘The merits of institutional research depend not on its scientific underpinnings, but on its relevance and influence in decision and policy making’’ (p. 35). That is not a call to abandon the canons of good social science and education research. It is recognizing that research is one compromise after another, and finding the balance between rigorous, thoughtful research and its practical and prudent application remains at the heart of both IR’s business and its challenge over the next decade. In being faithful to its role, IR professionals must not only be talented analysts (Tier 1) who understand the higher education landscape (Tier 2), but who are also thoughtful professionals who have absorbed and learned how to apply that understanding and contribute to effective decision-making on one’s own campus. I suspect few of us ever thought of T. S. Eliot as an institutional researcher, but the former English teacher in me thinks a quote from Eliot captures something close to the essence of what I believe institutional researchers are about. In ‘‘Little Gidding,’’ from ‘‘Four Quartets,’’ Eliot (1943, p. 59) wrote: We shall not cease from exploration. And the end of all our exploring Will be to arrive where we started And know the place for the first time. Thank you all very much for your patience and attention and for the honor of speaking this morning.

As an evidence-based practice, institutional research (IR) within higher education institutions involves data collection, analysis, interpretation, and dissemination to support policy decision-making (Webber, 2018). In many countries, recognition of the importance of this approach has led to the establishment of institutional research offices in postsecondary institutions. IR is a well-established practice in higher education systems in the United States, many European countries, Canada, and Australia, as well as in many developing countries. Taiwan is one of the representative nations in East Asian countries that has recently taken a leap forward in the field of institutional research (referred to as IR) in postsecondary education and has become more data-driven like in the United States. To date, influenced by the Ministry of Education’s (MOE) 2015 policy on Professionalization on Institutional Research, more than 70 % of Taiwan’s postsecondary institutions have set up Offices of Institutional Research (referred to as OIRs) in Taiwan (Lin et al., 2021). Some institutions assign additional functions to their administrative units, resulting in compound unit names such as Office of Institutional Research and Social Responsibility, Center for Innovative Teaching and Institutional Research, and so on. In addition, in schools promoting IR activities, there may not be a dedicated IR unit established due to the lack of manpower considerations; instead, IR functions may be incorporated into existing units (Lin & Chi, 2023). The most common setup is a dedicated the office of institutional research with a small group of staff, typically between 2 and 4 (Lin & Chi, 2023). The more robust and well-organized the OIRs are, the better they can establish a data-driven decision-making culture. Since 2018, Institutional Research (IR) has been integrated with the Higher Education Sprout Project (HESP) for colleges and universities to ensure university accountability and quality assurance. HESP aims to develop features of universities, nurture talents for the new generation, and enhancing student learning outcomes, as well as increase students’ competitiveness in the workforce beyond graduation (Ministry of Education, 2018). IR is an integral part of deepening higher education policy within the HESP to formulate direction for institutional development and strategic planning. Although IR has the potential to play an important role in the educational reform policy by providing evidence-based decision-making in higher education. Researchers (Fu et al., 2021; Lin et al., 2021) have pointed out that the greatest challenges faced by IR include high turnover rates, unclear researcher promotion systems, and a lack of direct training for institutional research talent in relevant academic departments within institutions. It was worth noting that the turnover rate among doctoral-level researchers is relatively high, and the average work experience is about 2-3 years. This is mainly because the ultimate goal of doctoral training in Taiwan often focuses on academic positions, Accordingly, when institutions cannot meet their promotion criteria, researchers with Ph.D. degrees tend to leave OIRs once they secure teaching positions elsewhere (Lin et al., 2021). Furthermore, directors of OIRs are often senior professors with concurrent roles as faculty within the university. Their terms are typically short, and they are also not familiar with the field of institutional research, leading to limited functionality of the office. In sum, if an IR unit experiences constant personnel turnover, its role and functions may gradually weaken due to changes in policies or key personnel. Consequently, it may be unable to provide a data-driven decision-making model for institutional development and sustainability. Therefore, a pressing issue is the retention of IR professionals, related to its status as a non-traditional profession operating in a “third space” (Calderon, 2018), characterized by the need for an interdisciplinary knowledge base and multi-disciplinary work-force, as well as constantly changing, context-dependent work roles (Whitchurch, 2008). As IR professionals provide intelligence for decision support (Calderon, 2018; Terenzini, 2013), building capacity to help institutions navigate through the change and uncertainty characteristic of the current era has become an IR priority (Leimer & Terkla, 2009; Volkwein et al., 2012). Nevertheless, the varied educational backgrounds and work histories of IR professionals, ecological frameworks of IR offices, and IR job-designs present a challenge to the development of uniform standards, codes of conduct, training, and clear career development paths, which are essential to expanding the IR workforce (Lin et al., 2018; Webber, 2018). Research has demonstrated that professional development is integral to expanding the IR workforce and building adequate IR capacity to support evidence- based decision-making. Among the current challenges to the development of IR as a profession, a main issue posing a threat to workers’ professional development efforts and IR capacity-building efforts is the shortage of skilled manpower resulting from high turnover rates and difficulties in attracting new talent into the profession (Calderon, 2018; Knight & Leimer, 2010; Lin et al., 2018). Furthermore, the current discourse on IR professional development focuses on services provided, skills required, and standards of practice (Ehrenberg, 2005; Terenzini, 2013; Volkwein et al., 2012; Webber, 2012), but little is known about such key elements of professional identity formation as how IR professionals acquire knowledge, develop professional practices, and internalize professional standards (Miscenko & Day, 2016). Professional identity represents the beliefs about, values related to, attitudes toward, and responsibilities of a particular profession that its members internalize, as well as the philosophy, ethical standards, and codes of conduct that they incorporate into their self-concept (Brown, 2015; Ibarra & Barbulescu, 2010; Miscenko & Day, 2016). Professional identity is thus an essential source of meaning in their work, influencing their mindset, acquisition of knowledge, development of competencies, and socialization into the behaviors and practices of the workplace. As such, it plays a key role in role fulfillment, job satisfaction and workforce retention (Ashforth & Schinoff, 2016; Brown, 2015; Miscenko & Day, 2016). Similar to other higher education support professions, IR work spans a broad third space between professional and academic domains, further complicating IR professional development, the meaning assigned to work roles, and the assessment of IR’s contribution to institutional progress (Calderon, 2018; Whitchurch, 2008). The literature suggests that the position, function, and context of third space professions may influence the construction and development of their members’ professional identities, highlighting the importance of understanding these processes in IR as an emerging profession. Social and narrative identity theories postulate that language plays a key role how people interpret and negotiate the meaning of their experiences as they co-construct and enact their identities (Ashforth & Harrison, 2014; Ashforth & Schinoff, 2016; Bamberg et al., 2011; Brown, 2015). Accordingly, to understand IR professionals formation of professional identities, we explored their stories of how they socially co-constructed the meaning of their work experiences. To guide this inquiry, we addressed the following research question: What do institutional research professionals’ narratives of how they conceptualize their experiences working in IR offices reveal about their professional identities and what shapes them? Literature Review Organizational Intelligence IR functions are key to sustainable change management (Volkwein et al., 2012). Terenzini (2013) asserts that, to fulfill this mission, IR professionals need to cultivate three core tiers of intelligence: (1) technical and analytical intelligence, (2) issues intelligence, and (3) context intelligence. Technical and analytical intelligence comprise knowledge of the data and information management structures of their respective institutions. It involves competence in accessing data and employing appropriate research methods and analytical tools to interpret the data (Terenzini, 2013). Issues intelligence entails having a comprehensive understanding of the structure, culture, and politics that underpin an institution’s operations and decision-making processes. Equally important is substantive knowledge of relevant literature to inform the analysis and interpretation of phenomena under study to generate tailored, evidence-based suggestions (Terenzini, 2013). Contextual intelligence encompasses nuanced understanding of the institution’s relationship to its external environment (local and global) and how it shapes the institution’s activities and functions. The relevant literature highlights the indispensability of these kinds of knowledge to IR’s capacity to reliably support accountability in decision-making and policy design, and reform at institutional and national levels (Ehrenberg, 2005; Volkwein et al., 2012; Webber, 2018). Nevertheless, tensions associated with the complexity, scope, and expectations of IR’s role complicate the development of the competencies and professionalism of IR professionals. Four Facets of IR Volkwein (1999), observing IR professionals’ constant engagement in negotiating the demands of different tasks, identified four faces of IR professionals: as information authorities, as policy analysts, as spin doctors, and as scholar/researchers. As information authorities, they monitor the state of their institutions and report to the stakeholders. As policy analysts, they evaluate and report on institutional progress and suggest internal policy reforms. As spin doctors, they balance the need to protect the institution’s image with the maintenance of accountability. As scholar/ researchers, they generate reliable intelligence for supporting the roles of the other three faces (Volkwein, 1999). Navigating complex roles and tasks can result in role ambiguity, conflict, and stress, resulting in employees’ frustration (Leimer & Terkla, 2009; Whitchurch, 2008), which contributes to the high attrition rates and shortage of human resources prevalent in the IR profession (Knight & Leimer, 2010; Lin et al., 2018). Positive professional identity can buffer the impact of role tension and enhance commitment to the profession (Brown, 2015; Miscenko & Day, 2016). Nevertheless, how IR professionals construct their professional identities remains unexplored (Volkwein et al., 2012). Professional Identities of IR This study is based on the premises of the process model of identification (Ashforth & Harrison, 2014; Ashforth et al., 2008; Ashforth & Schinoff, 2016) and the theory of situational learning in communities of practice (Lave & Wenger, 1991; Wenger, 1999), which capture the relational, dynamic and continuous nature of professional identity construction. Drawing from the perspectives of these theories, we first postulated that development of professionalism is at the center of IR capacity building (Leimer & Terkla, 2009; Webber, 2018). In addition, we proposed that professional identity is a vehicle by which professional standards are internalized and enacted in real practice. Professional identity, a subset of social identity, is formed through social interaction among community members and engagement in collective meaning-making (Ashforth & Harrison, 2014; Ashforth et al., 2008; Ashforth & Schinoff, 2016). By participating in both professional and social activities within and across organizations, professionals gain a clear self-concept of who they are and what is expected of them in the culture of their profession, which they internalize by mirroring the thought processes, feelings and behaviors exhibited by colleagues (Ashforth & Schinoff, 2016; Wenger, 1999). This process is called identification. During this iterative sense making process, members progressively exhibit beliefs, values, and behaviors consistent with their profession (Ashforth et al., 2008; Lave & Wenger, 1991). Thus, identity enactment and validation happen in tandem as other members affirm these dispositions and behaviors (Ashforth & Harrison, 2014; Ashforth et al., 2008; Ashforth & Schinoff, 2016). This reinforcement fosters a sense of self-efficacy, belonging, and professional commitment (Lave & Wenger, 1991; Wenger, 1999). In addition to identification, sense-making, and identity enactment and validation, members acquire language practices commonly used within their profession to articulate, communicate, and negotiate the meanings of relevant experiences (Ashforth et al., 2008; Ashforth & Schinoff, 2016; Bamberg et al., 2011; Ibarra & Barbulescu, 2010). This language allows professionals to define themselves based on their understanding of who they are, how they think others perceive them, and ultimately how they want to be perceived (Bamberg et al., 2011). Building on these assumptions, we argue that the embodiment and enactment of professional identities are contingent on the socialization process and context of the organization in which the individual is embedded. Accordingly, we adopted Ashforth and Schinoff’s (2016) conceptualization of the identity construction process (Figure 1), which provides an overview of the affordances for the sense-making. Sense-breaking, and sense-giving processes in the work context that underpin professional identities. We posit that identities are reinforced when individuals’ narratives of their past, present and future are validated within their social context. The language people use to talk about their work and lived experiences may be used as a lens to study identity (Bamberg et al., 2011; Ibarra & Barbulescu, 2010). Accordingly, based on the assumption that analyzing the language professionals use can shed light on how they continuously interpret and integrate their experiences to construct their professional identities, we investigated what Taiwanese IR professionals’ narratives reveal about how they conceptualize their work experiences and what shapes their professional identities. While IR practices have been successfully established in Taiwan, sustaining their expansion may prove challenging (Lin et al., 2018), which suggests the importance of examining the roles and contributions of IR work from the perspectives of those who provide it in order to support them and sustain the creation of value by IR functions in Taiwan. Study Design To conduct a cross-sectional study of how IR professional identity construction is negotiated, we adopted a constructivist perspective and qualitative research approach. As identity relies heavily on social interaction and language (Brown, 2015), we proposed that the constructivist ontology and epistemology, which assume that reality is subjective and meaning is co-constructed (Lincoln et al., 2011), are compatible with Bamberg et al.’s (2011) theoretical assumptions that identities are co-constructed through language and that that people’s social realities can be understood via their first-hand accounts of their experiences. Participants Using purposeful sampling, we recruited 11 IR professionals, six males and five females, currently working in IR offices across Taiwan, who voluntarily participated without external incentives. According to the survey results from Taiwan’s Association for Institutional Research (referred to as TAIR) (TAIR, 2021), approximately 40% of Ph. D holders work as principal researchers and are involved in data analysis as well as the discussions with institutional leaders in support with decision- making for institutional improvement and strategic planning. Since the nature of IR is interdisciplinary and more closely related to social science majors (e.g., higher education major), this study targeted those who are currently working or have worked in the IR office with a Ph.D. degree as our participants, who can better reflect the dynamic workplace of IR. Therefore, our selection criteria included holding a PhD, based on previous research indicating that PhD holders in non-academic positions often have extensive knowledge of the operations of the system, exhibit high commitment to their work, and have a strong sense of identity with higher education communities. Our sample comprised one IT specialist, six faculty members, and four IR staff members (see Table 1). Most of them are from the major of education, followed by business and informatics. Participants gave written consent by responding to our email detailing the project’s purpose, procedures, interview questions, data protection measures, and dissemination plan. Data Collection We developed and piloted an interview guide focusing on the scope of IR work; its relevance or value, workload, and support; and the nature of interactions in the participants’ immediate and national-level communities of practice (See Appendix A). Because we believed that identity narratives require deep reflection and sense making, processes in which the language of both respondent and interviewer plays an important role (Bamberg et al., 2011; Brown, 2015), we engaged participants so as to encourage them to converse using the language with which they were most comfortable. Since the interviewer was familiar with both Chinese and English, the interview process was conducted mainly in Chinese and English, but also mixed with English. The 40- to 60-minute interviews were audio-recorded and transcribed verbatim. It should be also noted that the difference in languages results in the loss of the original intended meaning in translation. However, we enhance the creditability of the data through triangulation such as participants’ confirmation with the transcripts, peer review and researcher’s reflection. Data Analysis Our analysis of the transcripts focused on three key points: (1) what it is like to be an IR professional, (2) what roles and competencies IR professionals associate with the IR profession, and (3) factors that influence IR professionals’ interpretations and appraisals of their work experiences. To analyze the data, we following Braun and Clarke’s (2006) thematic analysis steps of familiarization, developing initial codes, generating themes, reviewing themes, and discerning patterns, first manually and then using Atlas.ti coding software., then results. We conducted the initial step of data familiarization manually by reading and commenting on the transcripts and sharing the results. Then for qualitative data analysis we used Atlas.ti software, which has been used in the social sciences to facilitate systematic organization, analysis and visualization of data, including coding, commenting, memoing, grouping codes, and creating networks of codes or themes (Friese, 2019). First, to further familiarize ourselves with the data (step 1), we used the software to create free reflexivity memos. To develop initial codes (step 2), we employed both inductive and deductive coding techniques of Atlas.ti (i.e., open coding, in vivo coding, coding by list, and quick coding). Inductive coding was exploratory while deductive coding involved the use of structural codes, using key terms from interview questions and prompts to segment the data (Braun et al., 2019). Our unit of analysis was an excerpt carrying a single idea or episode reflecting a work experience. We focused on what participants said and how, looking for patterns across participants’ stories (Braun & Clarke, 2006). Using Atlas.ti’s comment and memo function, we were able to write more detailed thoughts about each participant’s transcript, including specific codes and quotes. The software then generated a list of codes and quotes, which we iteratively examined for patterns. We grouped codes that reflected similar concepts identified through the software’s query and group codes function into semantic categories, which we then clustered into broad themes. After reviewing the initial codes and preliminary themes, using the software’s network and code-merging functions to visualize the data and revise codes, we constructed and reconstructed categories in order to identify the dominant themes and sub-themes found in participants’ discourse about their profession. This iterative process of coding and collective reflection on themes continued until we reached theoretical saturation. Reflexivity Narrative research relies on researchers’ perspectives and experiences as essential tools for unpacking the meaning embedded in participants’ stories (Braun et al., 2019; Lincoln et al., 2011). We all have qualitative research backgrounds and expertise related to IR and professional identity. To enhance the integrity of our research, we engaged in the process of reflexivity throughout the study, which included keeping reflexivity journals, in which we recorded our own experiences and understandings of the concept of professional identity and of the target population. This journal- keeping helped us explicate our biases and keep them from influencing how we conducted the study. At the study’s outset, our first step to obtain a non-biased, realistic view of the current state of the organizational structures, roles and practice styles of IR in Taiwan’s higher education was to sample IR professionals across organizations in different regions of the island, taking into consideration their roles, tenure and gender. To gather rich accounts of professional experiences, we recruited only volunteers willing to participate without compensation, to whom we pledged to protect their real identities. To enhance the rigor of our research, one of our authors, who is not involved in IR, conducted all the interviews and the initial analysis, using inductive coding. Her outsider role contributed an objective perspective to the study. These measures, combined with the flexibility of semi-structured interviews and the initial use of inductive cod, helped us elicit fresh insights into the topic and minimized bias stemming from preconceived notions about IR. Subsequent analysis involved iterative cycles of interpretation, re-interpretation, construction-reconstructions of categories, and emergence of themes. Finally, we presented the preliminary findings at an international conference, where IR professionals provided positive feedback on the authenticity of the study’s themes and how closely they reflected their lived experiences. This affirmation of the relatability of our synthesis to their experiences confirmed the validity and meaningfulness of the findings. Results Two key themes with various sub-themes were generated (1) facets of institutional researchers’ professional identities and their related tensions, and (2) suggestions for developing the next generation of IR professionals. Tables 2, 3, 4, 5, 6, 7 provide illustrative quotes for the sub-themes along with informants’ IDs and genders (e.g., P1, Male) to provide context (see Table 1 for participants’ background information). Theme 1: Facets of Institutional Researchers’ Professional Identities and Related Tensions Our analysis of participants’ perceptions of themselves and how others perceived them suggested the narratives of their professional identities, which included coordinator, data analyst-scientist, sage-advisor, and scholar, though their conceptualizations of these identities were not without ambiguity and complexity. Their stories illustrated how discrepancies between their perceptions of themselves, their roles, and what was expected from them could cause tension. The narratives of these facets and their tensions are discussed below. Coordinator Narrative The participants’ work-related stories richly expressed the importance they placed on creating alliances and building relationships with offices outside IR. In the coordinator narrative, a good IR professional is positioned as an effective communicator (Table 2, quotes 1), which some participants’ contrasted with the disposition of the detached introvert, who may struggle to coordinate with other offices or communicate with stakeholders (Table 2, quotes 2 and 3). The Data Analyst/Scientist Narrative This narrative represents institutional research professionals’ strong belief that being analytically inclined is essential to thrive in IR work. This belief permeated their stories about the importance of data driven decisions (Table 3, quote 1) and sensitivity to data as it pertains to IR issues (Table 3, quote 2). They also highlighted the importance of the ability to present data to diverse audiences, which presented the challenge of managing varying levels of sophistication in reporting their analyzes and accommodating the expectations of different audiences (Table 3, quotes 3). The Sage/Advisor Narrative This narrative portrays institutional research professionals as consultants and advisors with the institutional intelligence to support decision-making in all matters related to the institution. Under this narrative’s umbrella are several roles in which IR professionals cast themselves: information resource manager, forecaster, and strategist outlining the institution’s future direction (Table 4, quote 1). Participants also expressed that embodying this identity entails risk as it puts one’s profession’s reputation on the line, especially when evidence-based intelligence or support is disregarded, manipulated, or misused to serve specific agendas (Table 4, quotes 2 and 3). The Scholar Narrative Most participants expressed how they valued opportunities for scholarship provided by some IR positions, which had a reciprocal relationship with either their professional development (Table 5, quote 1) or other positions that they held outside the IR office (Table 5, quote 2). However, the discourse also revealed how such factors as the imbalance between administrative and research responsibilities and certain policies on data usage (Table 5, quote 3) hindered their pursuit of scholarly development. Theme 2: Suggestions for Developing the Next Generation of IR Professionals When envisioning the future of IR in Taiwan, a majority of the participants focused on the development of the next generation of institutional researchers. Two key priorities permeated participants’ perspectives on the long-term sustainability of the IR profession in Taiwan: (1) strategies for retaining the current workforce and (2) strategies for attracting new people from various backgrounds. Suggestions for Retaining the Current Workforce Participants’ suggestions on how institutions can retain their current IR workforce and avoid brain drain included establishing clear and long-term career development pathways (Table 6, quote 1), improving professional development opportunities, and prioritizing specialization of roles in job-design (Table 6, quote 2). They strongly emphasized fully professionalizing IR offices by having full rather than part-time leadership positions (Table 6, quote, 2), formally recognizing employees’ valuable contributions (Table 6, quote 3), providing support for employees’ needs at all levels, and having viable long-term plans for IR development (Table 6, quote 4). Suggestions for Attracting New People Participants’ stories also focused on developing strategies for attracting and retaining new talent to the field. For example, particularly in the Taiwanese context, fostering a symbiotic relationship between faculty members’ own research agendas and their IR roles, could attract them to IR (Table 7, quote 1). Similarly, they recommended tailoring incentives not only to faculty members (Table 7, quote 2) but other types of IR professionals, such as IT specialists (Table 7, quote 3) and junior professionals (Table 7, quote 4). Discussion Our analysis revealed that these multiple professional identities and their tensions are intertwined. For instance, in the coordinator narrative, a good IR professional is positioned as an effective communicator with strong interpersonal skills to facilitate working in multidisciplinary teams. This expectation is characteristic of third space professionals, whose roles bridge professional and academic domains (Whitchurch, 2008). Terenzini (2013) emphasized the importance for IR professionals to build networks of relationships and collaborate with other offices, especially academic units, to maintain current issues-intelligence and increase other units’ reliance on IR services for decision support (Webber, 2012, 2018). Effective collaboration with other units requires a broad base of knowledge of key topics, which is integral to developing issues-intelligence (Terenzini, 2013). Also, participants’ narratives revealed that while they acknowledged that data analysis for decision-making and official reporting purposes took precedence over publishing, scholarly activity and development were important to them because in Taiwan, most PhD holders are trained to pursue academic careers. Further, the high turnover of IR professionals in Taiwan could be attributed to the lack of promising career pathways in the field and perceptions of the heavy administrative workload as a barrier to enacting one’s scholar identity. This finding supports previous claims that imbalances between administrative responsibilities, which are prioritized in the IR work cause role conflict (Ehrenberg, 2005; Knight & Leimer, 2010). The data analyst/scientist narrative highlights IR professionals’ role in shaping policies. Participants emphasized that acquiring good data, seeing connections in the data, and drawing on a broad knowledge base to interpret what they find were key to understanding higher education issues and devising effective policies. Another key is being versatile in the use of various analysis methods and able to present findings to diverse audiences. These requirements mirror Terenzini’s (2013) three tiers of technical/analytical, issues, and contextual intelligence for IR capacity building. In Taiwan, a main source of tension in the enactment of this identity is the lack of good quality databases for conducting analyzes that strongly support educational policymaking. This issue corroborates claims in previous studies that the validity and reliability of IR intelligence rely access to rich and high quality databases across disciplines (Ehrenberg, 2005; Lin et al., 2018). Therefore, we advocate for the merging of databases at institutional and national levels as an important goal for the IR profession. The significance of the three tiers of intelligence to IR professionals’ provision of valuable decision support is apparent in the participants’ sage/advisor narratives. These narratives highlighted their role as consultants who help other units analyze and evaluate their current initiatives and resolve strategic planning issues, for example, by as redefining their priorities for effective execution. However, participants’ narratives also made it explicit that the lack of quality databases hindered the enactment of the sage/advisor identity. Other difficulties in enacting this identity stem from the politically charged nature of higher education governance. To date, there is the lack of controls for ensuring accurate and legitimate use of the evidence produced by IR offices. Moreover, the analyzes and interpretations provided may unintentionally put other units under scrutiny, hence increasing the likelihood of tension with stakeholders in these units. Given this risk, IR professionals also need to develop skill in navigating a politically charged institutional landscape while providing valid interpretation of the data available for analysis (Webber, 2012, 2018). Ehrenberg (2005) observed that it is harder to convince decision makers who are not “data driven” than those who are to base their decisions on empirical evidence. Our participants similarly claimed that such disregard of data diminishes the IR’s contribution to institutions and compromises its position. Given that the literature documents that providing evidence for decision support is the cornerstone of IR work (Lin et al., 2018), we argue that this issue is a potential threat to the IR profession and its members’ professional identity. IR professionals’ embodiment and enactment of the coordinator, data analyst/ scientist, sage/advisor, and scholar facets of professional identity enabled them to fulfill their main responsibilities of assembling, analyzing, interpreting, and reporting institutional data. In turn, succeeding in these responsibilities depends on the communication, collaboration, data management, consultation, analytical and research competencies they develop through daily practice and interactions with others (Ashforth & Harrison, 2014; Lave & Wenger, 1991; Wenger, 1999). Our finding of the importance of having multiple identities which prioritize different competencies echoes previous research showing that IR is a unique amalgamation of particular skills and intelligences (Terenzini, 2013; Volkwein, 1999). Overall, our findings on facets of institutional researchers’ professional identities and their related tensions add to the current discourse on IR professionalism, which highlights the complex and dynamic nature of IR’s efforts to meet competing demands to support their institutions’ effective decision making. These demands often leave IR professionals struggling to meet all obligations of the roles that they have to play (Volkwein, 1999). Therefore, to help IR professionals navigate their complex work situations and meet their role expectations we recommend that institutions provide support through continuous professional development opportunities (Calderon, 2018; Webber, 2018) and recognition of IR professional identities. High attrition rates due to brain drain and lack of incoming talent were likely to undermine the contributions IR offices made to their institutions and were thus major obstacles to IR capacity building efforts. Therefore, our participants focused on the importance of ways to retain the current workforce and ultimately expand it. Toward this end they suggested such retention and expansion strategies as developing clear promotion paths, prioritizing full-time over part-time appointment, designing specialized roles, providing training programs, and accommodating the work needs of different types of IR professionals. These suggested improvements alluded to the importance of job security, career progression, recognition, belongingness, guidance, and work fulfillment as necessary factors for strong commitment to the IR profession (Brown, 2015; Ibarra & Barbulescu, 2010). Thus they could help sustain the professional identity and capacity development of the IR workforce. Combining insights from the two themes emerging from this study, (1) institutional researchers’ professional identities and their tensions, and (2) developing the next generation of IR professionals, reveals that human resources practices underpin the enactment, embodiment, and development of IR professional identities, and that organizational context and work culture play critical roles in these processes (Miscenko & Day, 2016), Thus, they largely influence whether these identities are weak or strong. The emerging literature indicates that organizational contexts afford employees different levels of belongingness, involvement and autonomy to explore all facets of their work, that is, different levels of empowerment (Ashforth & Harrison, 2014; Brown, 2015; Whitchurch, 2008). In particular, the socialization practices in the institutional environment such as interactions with colleagues and leaders, professional development, and explicit promotion paths affect the formation of employees’ professional identity (Ashforth & Harrison, 2014; Ashforth & Schinoff, 2016; Ibarra & Barbulescu, 2010). Furthermore, the participants’ work experiences and suggestions affirmed the importance of receiving support from their institutions and administrators as well as professional associations. Participants found working with a supportive director beneficial for handling work-related stress as well as experiencing empowerment, sense of belonging, overall work well-being and professional development. By offering nuanced insights into the importance of a fully committed, supportive director, our analysis expands on Knight and Leimer’s (2010) finding that perceived organizational support determines IR professionals’ intentions to leave or stay. As one participant declared, part-time supervisors of the unit are not conducive to sustainable development in the IR workforce. We argue, therefore, that to address the urgent issue of how to develop and maintain a productive IR workforce, any organizational practices undertaken to alleviate role ambiguity, role conflict, and role strain should start with effective leadership training for IR directors. Such programs should focus on helping directors develop ways to create a culture conducive to smooth operations that are based on understanding employees’ work needs, proper design of jobs, judicious deployment of workload, clarification of complex roles, and effective management of role conflict. This argument is supported by previous studies providing evidence of the association between leadership and employees’ productivity, job satisfaction, organizational commitment, and professional identities (Ashforth & Harrison, 2014; Knight & Leimer, 2010). Research Limitations This study has limitations that should be considered when interpreting the results. As it is a national study, it only samples IR professionals with doctoral degrees in Taiwan. While this is sufficient to establish credibility for studying various facets of IR professional identities in Taiwan’s higher education, future research should consider including IR directors and faculty members to utilize larger sample datasets and provide more generalizable information about workforce empowerment and career pathways for IR professionals in international higher education, Furthermore, future studies should employ survey methods to understand non-qualitative influences on IR professional identities and workforce empowerment in different types of higher education institutions. Longitudinal designs would offer more representative and deeper insights into how their professional identity develops over time. Conclusion Despite these limitations, our findings contribute to the discourse on the development of an IR workforce by highlighting narratives that represent authentic individual experiences, which are drawn together to illustrate a collective co-construction of what it means to be an IR professional. This construction is drawn from a set of common realities that Taiwanese IR professionals report facing in their respective work contexts. Similar to Volkwein’s (2012) IR roles of information authority, policy analyst, spin doctor, scholar/researcher, and knowledge manager, our participants’ narratives reveal multiple and interrelated identities as communicators, data scientists, consultants, and scholars. These findings provide insights into their vision for the future development of IR in Taiwan. To the best of our knowledge, this is the first study in which the professional identities of IR professionals at a national level in an Asian context are examined. Policy Implications As noted in the framework of Ashforth & Schinoff (2016), IR professional identity evolves naturally whether the individual accepts the role or not. Our findings indicate that the professional development of the IR workforce and the strength of its members professional identities can be scaffolded by supportive leadership and work environments to reinforce their professional identity for sustainable development in the IR workforce. Therefore, we suggest that teaching current leaders how to implement practices that empower their employees to meet complex role expectations can generate a chain effect across institutions as well as provide models for future leaders. This effect can contribute to countrywide efforts to provide the support for professional identity development, expand the IR workforce, and sustain IR’s capacity. Furthermore, our findings shed light on the importance of workforce empowerment for the construction of healthy professional identities and its dependence on perceived supportiveness of the workplace. This relationship has important implications for research on IR organizational cultures and workforce issues across institutions and countries. Such research can open avenues for reconceptualizing and redefining human resource practices that are suitable for reinforcing healthy professional identities. Therefore, we suggest that identifying factors that either empower or disempower IR professionals in fulfilling their work roles and enacting their professional identities need to be the target of future research on IR capacity building.

All is Far From Quiet on the Western Front Too often the climate between research administrators, ethics committee members and researchers is acrimonious. It seems there is an emerging culture of distrust between researchers and research ethics committees. The regulation of ethical conduct in human research has become something of an industry (Ashcroft 1999, p14), where a proliferation of rules has created an ‘out of control bureaucracy’ that often impedes the conduct of research (Haggerty 2004, pp. 392–394). It is not difficult to find examples of researcher frustration with the behaviour of ethics committees (Furedi 2002) or questions about the sustainability of the existing ethical review systems (Komesaroff 2002). Furthermore, there are repeated suggestions that the functioning of ethics committees may actually encourage an abrogation of the responsibility of researchers to reflect upon the ethical issues associated with their own research (Loff and Black 2004; Schwartz 2000). There are also warnings that the existing approach to the ethical review of research may actually be hampering (or ‘distorting’) useful research (Israel 2004, p. 6; Hoonaard et al. 2004, p. 11). Unfortunately, national research ethics frameworks can often be treated by ethics committees, administrators and institutions as though they were to be read and applied literally without a reflection on the specifics of a particular individual case. This can encourage a mindset that if an approach to an ethical dilemma is not explicitly anticipated in the national guidelines it is not permissible. It also tends to result in specific provisions of the guidelines being applied in a one-size-fits-all manner, where the realities and context of the specific project are ignored. Experience suggests that this kind of situation most frequently occurs when an over-worked ethics committee is confronted with a research design, approach or situation which is unfamiliar. It is important to remember that, in many jurisdictions, responsibility for the conduct of researchers, and the ability to take action in response to inappropriate conduct, resides at the institutional level. An important first step in the improving the institutional climate between researchers and research ethics committees is recognising that the ethical conduct of researchers is an institutional governance issue, which should be imbedded within the institution’s broader governance framework. Institutional Risk Concurrently with the concerns about the operation of research ethics systems, ‘regulators’ are placing an increasing focus on the responsibilities of institutions to ensure the ethical conduct of research conducted within the institution (Breen 2002). In many jurisdictions institutions must ensure researchers are trained in their ethical obligations. Institutions must also have systems to ensure compliance with the national standards and to actively monitor the conduct of researchers. A failure to meet these institutional responsibilities can attract serious penalties – for example all research funding1 to the institution may be suspended (for discussion about the ‘Johns Hopkins case’, see Abate and Russel 2001; Hotchin 2001; Keiger and de Pasquale 2002). Indeed, there are some very public examples of an entire institution being penalised for the actions of a single research team and/or because of perceived systemic failures within the institution. This has resulted in a climate where institutions feel unable to trust researchers to ‘do the right thing’. Instead, many institutions turn to a bureaucratic structure to ensure compliance with ethical and regulatory standards (ALRC and AHEC 2001, pp. 109–110; Dodds et al. 1 The research funding to an institution comes from multiple sources, so any impact upon funding (e.g. from the National Health and Medical Research Council) might initially impact upon a section of research funding to the institution. However, given the degree to which multiple funding bodies require institutional compliance with the relevant national framework, there may be a cascading effect that impacts upon most, or at least a significant proportion of an institution’s funding. 1994, p. 19). This in turn perpetuates an adversarial tone to the interactions between researchers and ethics committees (Shaul 2002). In such a context, ethics committees are perceived as having a ‘policing’ role, with the objective of catching researchers in wrongdoing. The combination of an institutional focus on compliance, with a view that researchers regard the work of HRECs with something between truculence, petulance or ignorance, appears to dominate the thinking of a majority of administrators, ethics committees and even regulators. This is the perceived ‘compliance problem’ that must be tackled – largely to safeguard the institution from the serious consequences of a regulator or funding body judging the institution to be non-compliant. The obvious limitation of this increasingly dominant way of viewing the interaction between researchers and research ethics committees is that it tends to be conceptually constraining. Within this constraint, any reflection on the role or operation of ethics committees tends to be almost exclusively focussed upon the goal of ensuring compliance, and compliance tends to be understood in terms of a researcher completing an application form for review by an ethics committee (Loff and Black 2004). In effect, institutions perceive the ‘compliance problem’ to be any researcher who does not submit their work for ethical review or otherwise adhere to the institution’s research ethics processes. However, in fact the ‘compliance problem’ might actually be the deleterious impact upon research practice when an institution relies upon directive rules and the threat of sanctions to modify researcher behaviour. The limited perspective of focussing on ethical review as a question of compliance ignores the interesting question of why a researcher, who might otherwise ascribe to the highest ethical standards and holds a very real personal commitment to the protection of research participants, might adamantly refuse to submit his work to review by a research ethics committee. It also glosses over the problem of a researcher who dutifully submits her work to ethical review, but in fact only begrudgingly pays lip service to the principles that supposedly underpin that review. As noted by Haggerty (2004, p. 410) a schism has emerged between the actual ethical conduct in research and the mechanism of following the rules. One of the reasons governing researcher practice can be so problematic is the realities of academic practice within a multi-disciplinary university, which can often be spread over geographically dispersed campuses. It is practically impossible for an institution to know exactly how individual researchers are conducting themselves. Given the degree to which inappropriate behaviour by even an individual researcher can expose the institution to risk, this creates understandable anxiety and a desire to try to centrally regulate conduct in a way that can be demonstrated to an external agency. Could it be that, we have allowed the goal of demonstrable compliance to subsume the principle of facilitating excellent ethical research and protecting the rights of research participants? It seems that in our rush to minimise institutional risk and to have a ‘sound’ governance framework, somehow research ethics has become about expertly filling out a form for a committee, rather than an ongoing active reflection on important issues and their application to useful practice. The Typical Approach to the ‘Compliance Problem’ In many cases, the typical governance approach to the ‘compliance problem’, seems to be predicated on an unquestioning acceptance of what Woodward (1994) described as ‘Weberian Orthodoxy’. This approach assumes that the best way to regulate practice within an organisation is through the release of detailed centrally formulated rules that are intended to closely direct practice in an absolute way. The rapidly increasing layers of rules and expanding role for ethics committees is a manifestation of this ‘Orthodoxy’ (Haggerty 2004, p.394). Also implicit in this dominant approach to research ethics systems is the use of what Hart described as a sanctions and enforcement model of control – where it is assumed that if the mere existence of the rules is not enough to modify behaviour then the threat of severe penalties will be (Hart 1999). A typical symptom of this orthodoxy, are ‘research ethics educational’ strategies that focus upon the correct completion of forms, the mechanism by which work is submitted for ethical review, and the textual features of informed consent materials. Often absent from such programs is much in the way of discussion about ethical conduct during the research, or resources to assist a reflective approach to the ethical challenges researchers might face. In fairness to the typical ethics committee, it should be acknowledged that a combination of crippling workload and limited resources can makes it difficult for a committee to do anything other than review the huge volume of new applications submitted to each meeting (McNeill et al. 1990; Federman; Hanna et al. 2003, pp. 8–9). There is simply not enough resources or time to enable the committee to be proactive in other areas (Breen et al. 2005, p. 469). This overwhelming workload is often cited as the reason a HREC does not take a more active role in educational strategies or policy development (Frew 2001, p. 67). But it could reasonably be asked, is the heavy workload of ethics committees and administrators really the source of the problem, or is it a symptom? As long as administrators and committee members continue to accept without question the notion that tackling the ‘compliance problem’ is a priority issue and the best way of tackling this problem is by modifying the behaviour/knowledge level of researchers through threats, ‘evangelical education’, or better ‘form filling’, then it might usefully be asked whether more time or resources would necessarily improve the climate between researchers and research ethics committees. Indeed there is a very real risk that greater resourcing for a research ethics committee’s compliance/educational strategies might in fact compound the problem. An educational program based solely upon the goal of reducing compliance problems by addressing ‘researcher ignorance of ethical standards’, might easily be construed as extremely patronising by researchers and further poison the relationship between committees and researchers. Limitations of the Typical Approach There are sources of evidence available that point to limitations of the typical approach to research ethics. Firstly, there continue to be public cases arising from breaches in ethical standards by researchers and/or failures in institutional research ethics systems (Unknown 2000; Oakes 2002 p450; Shaul 2002). Furthermore, some analyses of research practice point to a worrying level of ethical problems (Dotterweich and Garrison 1998, p. 444; Payne 2000, pp. 2–7; Thompson 1984). Secondly, there is a significant body of commentary from researchers that point to serious concerns about the approach to the application of research ethics systems (Fitzgerald and Yule 2004, pp. 36–38; Oakes 2002 pp. 445–446; Singer 1989). Thirdly, there are worrying indications that research designs are being compromised or ‘disfigured’ by existing ethical review processes (Haggerty 2004, p. 412; Hoonaard 2004, pp. 11–12; Iphofen 2004), or that researchers may actually be presenting a distorted description of their research to get it past an ethics committee (Shore and West 2005). Lastly, the simple fact that there is such an energetic international discussion about the sustainability of research ethics systems (Ramcharan and Cutcliffe 2001; Dodds 2002; Fitzgerald and Yule 2004) serves as indication that there is a level of institutional disquiet about the appropriateness and results of the current approach. Problems in the Breach It is perhaps tempting for research administrators and ethics committee members to blame any problems with the typical approach to research ethics on our respective national frameworks. Certainly, in Australia, the current edition of the National Statement on Ethical Conduct in Research Involving Humans (NHMRC 1999) is not without its limitations and short comings. The National Statement cites biomedical texts, such as the Declaration of Helsinki as the source of its authority. At the same time the National Statement fails to acknowledge other sources of ethical guidance for researchers outside of the health sciences, such as codes of conduct for psychologists (Dodds et al. 1994, p. 16). Much of the language and specific matters addressed within the National Statement are focussed on biomedical issues, with apparent assumptions about research practice that are more relevant to quantitative approaches. Fortunately, it appears that the next edition of the National Statement will address many of these issues and be far more relevant to the practice of research outside of the health sciences. Despite the temptation to blame our current difficulties on the relevant national guidelines, it should be acknowledged that, even when a national framework articulates principles with broad application to all human research, the way in which they can be applied at the institutional level is the source of many of the difficulties. In his report for the New South Wales Bureau of Crime Statistics and Research, Israel (2004) quotes specific cases of how the sometimes bizarre demands of an ethics committee can seriously impact upon the conduct of criminological research. This situation results in delays, considerable frustration and distrust, and can result in research being distorted to the point where it is unworkable. In these cases, the problem was not whether there was dispute over the validity of an ethical principle, but instead related to a policy position and approach adopted by an ethics committee. My own experience during the last ten years working with ethics committees in Australia, Canada, England and Vietnam is that committee members often feel isolated from, and under appreciated by, their own institution. This isolation can lead to a sense that members are charged with “protecting the institution”, but are unsure whether the institution supports them. When combined with a committee’s phenomenal workload (Shaul 2002; Allen 2005, p. 3), this insecurity and isolation can often result in a committee taking what Haggerty (2004, p. 412) described as an overly cautious and conservative positions on issues – even when the national framework enables them the freedom to assess the specifics of a case and consider alternative approaches (Haggerty 2004, p. 412). The approach to multi-site research is a good example of where ‘failures’ in the current approach to research ethics can occur at the institutional, rather than at the national level. This is an area that is a common source of criticism of human research systems. Researchers frequently complain about the inefficiencies, delays and frustrations associated with having to apply to multiple ethics committees for one study2. This occurs in Australia despite the fact that the current edition of the National Statement urges institutions to “minimise unnecessary duplication in review of multi-centre research” and suggests an ethics committee can decide to “accept the decision of another institution, organisation or HREC in relation to multi-centre research” National Statement (1999). Gold and Dewa (2005) point to similar problems with the processing of multi-site research in the USA. Like Australia, the American system allows for, and even actively encourages, ethics committees to accept the review decision of another committee for multi-site research (NBAC 2001, p. 122). In each case, the problems and frustrations faced by researchers conducting work at multiple sites has very little to do with the relevant national research ethics framework, but a great deal to do with the approach at the institutional level. It seems there is an active resistance of the use of reciprocal or centralised review solutions – even when the national frameworks encourages such an approach (Lux et al. 2000, p. 1183). When discussing this issue with ethic committee members and administrators I have frequently heard comments justifying this kind of resistance on the basis of the committee’s duty of care, responsibility and culpability. Such comments are often wrapped in deceptively compelling references to the law and insurance. Yet the implicit principles that underpin such arguments rarely seem to have been tested with the institution’s legal area. Furthermore, as noted above, any refusal to engage in some form of reciprocal or fasttracked review is contrary to the specific provisions of the relevant national framework. An institution involved in a multi-site project, does indeed have its own governance responsibilities for that research –including a duty of care to staff, students or patients. However, the presence of such responsibilities does not create a requirement for review by a research ethics committee. It does require that the institution have appropriate governance systems to ensure that the risks and issues are appropriately assessed and addressed. Rather than leaving ethics committees to struggle with the question of their responsibility, duty of care and accountability, there is a need for an institutional-level policy decision about how to handle multi-site research, including issues such as reciprocity. Facilitation and Legitimacy as a New Theoretical Model for Governance Systems If we are to tackle the goals of managing institutional risk, supporting ethical conduct in research, and protecting the interests of research participants then what is required is a complete departure from the existing approach to research ethics. No longer should we allow research ethics, and research ethics committees to operate independently of an institution’s broad approach to research governance. No longer should we allow ‘doing research ethics’ to mean filling out a form and interacting with an overworked committee that is struggling to work out its place within the institution. For local research ethics arrangements to have any hope of having a positive impact upon the design and conduct of research, they must first be predicated on the goal of facilitating excellent and ethical research. 2 For example, in 2005 this matter received considerable attention during the plenary and concurrent sessions of both the NHMRC conference on human research ethics and the annual conference of the Australasian Research Managers Society. I suspect that very few research administrators or research ethics committee members would disagree that facilitating research is important. Indeed, in Australia the National Statement (NHMRC 1999, p. 1) specifies that, in addition to its primary purpose of safeguarding the rights and welfare of participants, “an important secondary purpose” is to facilitate research. However, ethics committee members and administrators need to embrace facilitation as a core element of the operation of an institutional research ethics system. Rather than trying to catch researchers in wrongdoing or ‘educate them in ethics’, the whole focus of the system becomes a collaborative exchange between the stakeholders. Of course, that is not to imply that the primary objective of safeguarding the welfare and rights of research participants should in any way be compromised. Embracing the tenet of facilitation just means that the various players in the system try to work collaboratively with researchers, rather than try to ‘police’ behaviour. In such an approach, there is an attempt to build an equal partnership between researchers and the ethics committee. In her book on trust in society Misztal (1996, pp. 26–27) discussed how trust plays an important role in the support for, and participation in, public policy. Misztal (1996, pp. 245–269) suggested that the degree to which individuals see any exercise of public power as legitimate is a fundamental factor in determining whether they will comply with that power. I would suggest that a strength of this perspective on the interaction between instruments of regulatory control and the behaviour of practitioners is that it recognises the realities of decision making and practice in what Lash (1994, pp. 208–209) referred to as post-modern organizations (Lash 1994, pp. 208–209). For example – researchers have a degree of autonomy, they operate within fluid working arrangements and teams, and face a dizzying array of challenges that defy any directive-rule-based attempt to regulate their practice. It is simply disingenuous to assume that a central prescriptive rule can offer applicable solutions to the ethical challenges that researcher will face. We need to recognise that, especially for researchers in some disciplines, there may be perceived legitimacy problems with the national research ethics framework. Consequently, institutions have an important role to play in filling this legitimacy gap through the implementation of local processes and practices that can be perceived as being more relevant and applicable to the practice of researchers within the institution. Rather than trying to use directive rules to prescribe conduct, an institution’s research ethics arrangements need to encourage reflective practice where the researcher has an active role in formulating, and justifying, the solution to the issue – not just at the point of seeking ethical clearance, but throughout the entire life of a research project. The Griffith Model Griffith University is a multi-campus university based in the southeast corner of Queensland, Australia. It is a vibrant and growing institution, within what is called the Innovative Research Universities Australia group. The situation in terms of the approach to human research ethics at Griffith University, prior to 2003, was not that different from the problems and difficulties faced by many other Australian universities. For a variety of historical and structural reasons, research ethics had become isolated from the University’s broader research governance framework. This situation compounded the tendency of the ethics committee to take a cautious and conservative approach to risks and unfamiliar research designs. Furthermore, an overwhelming workload and insufficient resourcing meant that much of the University’s guidelines had not been updated for a number of years and predate the release of the current national research ethics framework. The negative impact of this situation was most obviously discernible in three ways: (1) The Director, Office for Research, the Deputy Vice-Chancellor (Research), and even the Vice-Chancellor, were receiving repeated complaints from senior researchers about “the ethics problem”; (2) even though there was a continual growth in the amount of research ethics activity, this was largely concentrated into a handful of broadly health-related disciplinary areas; and (3) there was increasingly an adversarial, even besieged, tone to the relationship between the HREC and the research community. In 2003, the then Deputy Vice-Chancellor (Research) commissioned a major review of human research ethics at Griffith University. The report of the Research Ethics Review 2003 was endorsed at the executive level. This provided top level support for a reform agenda. A new model for the governance of ethical conduct of human research arose from this review. The key features of this new model are outlined below. Executive Officer Support The University created a new senior policy officer position, at equivalent to senior lecturer level. Separate from the traditional secretariat support for the ethics committee, this position performs an advisory, policy development and educational role. Furthermore, there has been a limited ‘buy-out of the Chair and Deputy Chairperson of the ethics committee. Research Ethics Advisor network Every academic element of the University has appointed one member of academic staff as a Research Ethics Advisor (REA) for their area. REAs undertake a number of roles including: providing local ‘judgement free’ and discipline relevant advice to students, supervisors and other researchers; delivering workshops in their area; participating in the expedited review system (see below); advising the head of the element on research ethics matters; and facilitating communication to and from the committee. Proportional Ethical Review The University now has three ethical review pathways, which tie the amount of paperwork, the rigour of the review; and the processing time to a smart assessment of the risks and ethical issues associated with the proposed work. Two of these review pathways are online, with an online tool to assist the applicant accurately determine the appropriate review pathway. Research Ethics Manual The University has produced an online, booklet-based manual for researchers. Each booklet is focussed on a particular ethical issue or research design (such as online research). Even though the manual does outline external regulatory requirements and approaches previously accepted by the ethics committee, these are not presented as directive rules. Instead the Manual is intended to inform a researcher’s understanding of the principles and to assist them to justify alternative approaches. Case Study Training The approach to research ethics training within the University is not based solely on the process of ethical review and better ‘form filling’. Instead, real cases are used to explore issues that can be raised during the design, ethical review, conduct and publishing of human research. This approach to training stresses a two-way learning process, where the experience of researchers is fed back into the institutional approach to governance. Targeted Arrangements The University has adopted special policies and processes for research undertaken by coursework and undergraduate students, multisite research and research involving ‘standard’ instruments. There are also supplementary processes for research involving Indigenous Australians, clinical trials, and research conducted overseas. The Griffith Experience: 2 Years On At the time of writing, the ‘Griffith Model’ has been operating for just over 2 years. To date, there has not been any systematic collection of data about the impacts of the ‘Griffith Model’. However, there are some sources of data and testimony that make it possible to a t least suggest the degree to which the changes to governance at Griffith University have had a significant and enduring positive impact. Attitudes Towards Research Ethics In 2005, the Office for Research at Griffith University was subject to a comprehensive external review by a panel of executive-level academics and senior research administrators. The review process involved the review panel meeting with a range of researchers and key stakeholders, as well as considering confidential submissions from across the University. In their written report noted (Glover et al. 2005): & That the institution’s research community showed significant support for the changes to the University human research ethics governance arrangements (p. 17); & That there was “uniformly strong endorsement” for the University’s research ethics system, which was described as “effective and responsive” (p. 8); and & That the work and contribution of the new ethics team was greatly appreciated (p. 8). Clearly this represents a significant turnaround in climate at Griffith University in only a few short years. This major change in climate is also reflected in my anecdotal observations and in the feedback that I continue to receive from researchers and support staff. Participation Rate In 2004/2005 a total of 559 applications for ethical review were submitted at Griffith University, as opposed to 430 applications in 2003/2004, representing a 30% increase in activity. This followed a 40% increase in activity 2003/2004, compared to 2002/2003. The significant increase in new applications in 2004/2005 came at the same time as there was a 150% increase in variations to approved projects. As part of the ‘Griffith Model’, new processes were implemented to make it easier for a research team to extend and vary an approved protocol to cover the next phase of a research project. In practice a variation can often negate the need for a research team to submit a new ethics application for a new phase of previously approved research. It is also worthwhile commenting on how the above activity is now spread across individual University elements. During the period 2003–2005 some areas of the University have seen massive increases in activity, such as: Environmental Planning; Film, Media and Culture; Hospitality, Sport and Leisure; Information Computer Technology; Marketing; and the Queensland Conservatorium of Music. These are areas that traditionally had relatively low levels of human research ethics activity. To some degree, these increases in activity can be attributed to the maturing research profile of Griffith University. However, clearly increases of the order reported above are not reflected in improvements in the University’s grant success rates, publication rates and research higher degree student completion rates. I would suggest that, to a large degree, the massive jump in activity reflects a surge in researcher participation in, and engagement with, the University’s governance framework. I would contend that the work to address the ‘legitimacy gap’ of the national framework and previous University arrangements has resulted in not only more researchers electing to participate in the framework, but also a more complete spread of this activity into disciplinary areas outside of the health sciences. Client Service, Efficiency and Workload In 2004/2005 almost 90% of Griffith University’s human research ethics activity was reviewed via one of the two expedited review pathways (Allen 2005, p. 5). Coupled with the variation mechanism, course clearances, approved protocol procedure and multi-site arrangements this made a considerable contribution to the way in which the needs of researchers (the clients of many of the systems) are met. In many cases, requests from researchers were resolved in a matter of a few days. The changes outlined to the University’s governance framework also generated considerable administrative and committee efficiencies. For example, rather than the ethics committee struggling through in excess of 30 applications at each of its monthly meetings, the committee now reviews an average of five applications. Rather than making 20+ copies of 30 applications, and posting these to each member, Office for Research needs only expend a fraction of the resources to support the new multi-pathway review process. A significant impact of this change is that the ethics committee now has time to actively reflect upon policy issues and discuss emergent issues. The committee can also now afford to take the time to carefully reflect upon the ethically sensitive research that is submitted for full review. It is unlikely that the previous University systems, resources and arrangements could have coped with the recent increases in workload. Even the modest growth in activity seen in the years prior to 2003 would probably have made the HREC’s workload unsustainable. During this period the University has established a new medical school, and commenced strategic collaborations with bodies such as the Queensland Institute of Medical Research. Despite the extra volume and complexity arising from these institutional developments they have not placed any significant strain on the governance framework or the HREC, because of the reforms commenced in 2003. Nevertheless, it would be extremely short-sighted to believe that further work is not needed at Griffith University, or the challenges ahead will not demand further innovation. Conclusion The typical approach to research ethics is neither successful or desirable. In many cases it generates a culture of distrust and resentment between researchers and research ethics committees. At worst, the typical approach may be distorting or blocking useful research, without necessarily having a positive impact upon research practice or the interests of research participants. Institutions have an important role to play, by establishing local arrangements that address any legitimacy gap between the relevant national research ethics framework and research practice. The focus of these arrangements needs to be on fostering and supporting reflective practice, not unquestioning rule following. The experience at Griffith University suggests that it is possible to have a significant impact upon the climate between researchers and ethics committees, improve client service and efficiency, and achieve better outcomes, from the adoption of an approach that treats research ethics as an institutional governance issue.

Graduate social work students must be provided the necessary knowledge and skills to become critical consumers and producers of research. The Council of Social Work Education, in their Educational Policy and Accreditation Standards (2004), identifies research as one of the core foundational curricular content areas. The Council describes the function of research by stating that: ‘‘Research knowledge is used by students to provide high-quality services; to initiate change; to improve practice, policy, and social service delivery; and to evaluate their own practice’’ (p. 10). This statement clearly indicates how research is integral to all realms of social work practice. The National Association of Social Workers (NASW) Code of Ethics (1999) further highlights the value of research by identifying it as one way to achieve the profession’s core value of social justice. The desire to ‘‘promote social justice and social change,’’ however, should not obscure the possibility that research can produce individualand societal-level harms. The code outlines ethical considerations such as, ‘‘Social workers engaged in evaluation or research should carefully consider possible consequences and should follow guidelines developed for the protection of evaluation and research participants. Appropriate institutional review boards should be consulted’’ (section 5.02[d]). Although the NASW code states that social workers ‘‘should’’ consult with the appropriate review board, in actuality university-based social work researchers must submit a human subjects application to an Institutional Review Board (IRB) when their projects meet the definition of research as specified by the federal regulations. The IRB process aims to strengthen research ethics and to assure that study participants are not exploited. The IRB and Student Projects Social work has a growing body of literature related to the IRB process (e.g., Melville, 2005; Shore & West, 2005; Sanders, 2003), but no articles that specifically look at the IRB process as it relates to student projects. Two nonsocial work articles were located that addressed this issue, one from public health nursing (Foss, 2005) and another from business (Jenson, Mackiewicz, & Riley, 2003). Both articles describe challenges and benefits with the IRB based upon the authors’ experiences. A primary identified challenge was the amount of time required to complete the IRB review. Other challenges identified by Jenson et al. included how the IRB process may not be fully appropriate for student projects, and compliance with the IRB increases research instructors’ workload. Both articles describe the benefits as including increased student awareness of ethical research, as well as improved research designs. This article addresses the gap in the social work literature regarding research instructors’ experiences guiding graduate students through the IRB process. The study examined what are the benefits and challenges of having students experience an IRB review, as well as the processes by which instructors determine whether a review is required. The study’s ultimate aim is to provide research instructors recommendations on how to facilitate the IRB process of student projects, with an eye towards promoting ethical research. Prior to presenting the study methodology and findings, a brief overview of the federal human subjects regulations and IRB process is provided. The Common Rule and the IRB The federal human subjects regulations (45CFR46), also known as the Common Rule, define research as ‘‘a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge’’ (United States Department of Health and Human Services, 2001, §46.102). Confusion persists over what constitutes research. The National Bioethics Advisory Commission (NBAC, 2001), for example, claims that the Common Rule’s definition does not fully distinguish between practice and research. As a consequence, NBAC writes, ‘‘there is wide variation among federal departments and agencies regarding their policies and procedures for determining whether a research activity is exempt from federal regulations’’ (p. 10). Jenson et al. (2003) describe how universities have varied interpretations of the regulatory definition resulting in different requirements for the review of student projects. More specifically, some institutions require IRB review while others view student projects as not contributing to generalizable knowledge and therefore falling outside IRB purview. The authority of the IRB, as specified by the Common Rule, includes approval, approval with contingencies, deferral, or no-approval of research proposals; continuing review; observing, monitoring, and auditing of research projects; and suspension or termination of approval. The charge of the IRB is to assess research proposals across the Belmont Report’s (1979) three ethical principles: respect for persons, beneficence, and justice. Respect for persons is actualized through the guidelines established around the informed consent process. Beneficence is actualized through a risk/benefit analysis. IRB members assess the principle of justice by determining whether the research benefits the community, avoids exploitation, includes all affected populations, and considers demographics as well as access issues (McGough, 2001). The Common Rule specifies criteria for three levels of review: exemption, minimal risk, and full review. An exempt project must meet one of the six criteria listed within the regulations. Studies eligible for minimal review are those in which risk is assessed to be no greater than what can be expected in daily life. Full review is required for a study where risks include potential damage to the physical or psychological health, reputation, or economic welfare of a subject. METHODOLOGY The sample included 16 social work graduate research instructors from different Council on Social Work Education (CSWE) accredited schools. Sample size of this exploratory study was determined by saturation guided by the intent to gain an in-depth understanding of these instructors’ experiences. Participants were identified primarily through Web-based searches and via the National Association of Deans and Directors listserv. Of note, study recruitment occurred only after the author’s University IRB reviewed and approved the study. The average length of time teaching as either an adjunct or full time social work faculty member was 12.4 years, with a range of 5 to 30 years. Fifty percent of the sample identified themselves as Assistant Professors. Other interviewees included Associate or Full Professors, Deans, and lecturers. Nine of the interviewees served on their University IRB. The average length of tenure on the IRB was 4.9 years, with a range of 2 to 8 years. Interviews were conducted over the telephone and lasted approximately 45 minutes. The primary interview questions this article reports upon included (a) what are the benefits and challenges of having student projects reviewed by an IRB, (b) how do you determine whether a project requires IRB review, and (c) what guidelines or recommendations do you have to facilitate the review of student projects? Construction of the interview guide was informed by the author’s experiences teaching research and from serving on IRBs, as well as from the literature. All participants consented to have the interview audio taped, which were transcribed verbatim. Constant comparative method guided data analysis (Lincoln & Guba, 1985). As an overview of the analysis, the verbatim transcripts were coded line-by-line, moving towards the development of provisional categories. These categories corresponded to the questions asked within the interview guide. Transcripts were then imported into QSR N6, a qualitative data software package. Each category was then examined to identify what subcategories emerged from the data. The categories and subcategories were compared both within a given transcript and across transcripts, which resulted in further reconsideration of the category structure. FINDINGS IRB Benefits Potential benefits of having student projects reviewed by the IRB primarily focused upon how an IRB review: (a) creates an awareness of the process and reinforces the importance of research ethics, (b) forces students to conceptualize their projects prior to study recruitment, and (c) provides an additional layer of protection for both the potential study participants as well as the instructor. Of note, for some of the interviewees the review is actually conducted by an internal school committee rather than the University IRB. Creating Awareness and Emphasizing the Importance of Ethical Reviews Students may gain a deeper understanding or appreciation of the importance of research ethics by having to submit an IRB application. The IRB process becomes real and not just a topic covered abstractly. By filling out the application, students learn what types of questions are posed by an IRB while having to struggle with the application of these questions to their own proposed projects. An ability to complete an IRB application benefits students who engage in research post-graduation. Even if students are not directly involved as producers of research, knowledge regarding the IRB likely increases their sensitivity to ethical considerations, prompting them to critically assess other researchers’ proposed projects. For example, students would have insight regarding appropriate informed consent content. Simply put, several interviewees described submission to the IRB as ‘‘part of the educational process.’’ The time spent on protection of human subjects also communicates the seriousness of ethical conduct. Or in the words of one interviewee, the IRB ‘‘isn’t just some little rinky-dink thing that they’re doing for an academic exercise to get a grade. That we take it seriously, and to make them consider the protection of the people whose information they’re taking.’’ Encourages Conceptualization By completing an IRB application, students are forced to conceptualize their projects prior to engaging in recruitment or data gathering efforts. As students quickly learn, an IRB proposal is likely to be rejected if the project is not thoroughly conceptualized. The IRB application and the consequent feedback prompt students to identify potential risks and strategies to minimize these risks. As an example, an interviewee described the benefits in the following manner: It helps them to organize their thinking, they look at their project in a different way versus oh I just have to do these surveys, you know, versus how can I get the information that I want while also not coercing [participants]. The process of completing an IRB application furthermore represents an opportunity to sharpen one’s writing skills given an expectation is the application is clearly presented. Adding a Layer of Protection Receiving IRB approval for student projects also provides a level of security for the instructors. One of the identified values of having an independent body review student proposals was the identification of risks that the student and the instructor may have overlooked. For some of the interviewees, IRB approval is comforting as it provides a sense of security knowing that an institutional entity would support the instructor in case something goes wrong with the student project. This protection is welcomed given today’s ‘‘litigious society.’’ IRB Challenges For some of interviewees, there were no identified challenges related to the IRB review of student projects. A lack of challenge was attributed in part to the availability of concise checklists that summarize the requirements of the IRB, which at times were coupled with a streamlined review process specific to student projects. Other interviewees, however, identified challenges that touched upon problems related to (a) the length of time required to prepare an IRB application and gain approval, (b) lack of IRB resources, (c) negative faculty perception of the IRB, and (d) IRB committee composition and dynamics. Additionally one interviewee summed up the challenges by stating, ‘‘the benefits are the challenges, you know in terms of pushing limits, the pressure, getting them to think [the project] through.’’ Time-Consuming A primary challenge is the amount of time required to prepare an IRB application and to receive approval. Students must first understand research methodology and the underlying ethical principles, then learn how to complete an IRB application, which entails learning how to conduct a risk/ benefit analysis and how to create informed consent documents. Instructors must factor in adequate time in their syllabi to cover all of these areas. Additionally instructors must allow time to guide students on how to respond to the IRB feedback and to monitor all of the projects to assure IRB compliance. Given that at some institutions IRBs meet relatively infrequently, students must be prepared to submit their proposals according to the IRB schedule. The research instructor is responsible to assure students are aware of the IRB schedule. Due to the time consuming nature of the review process, one interviewee reflected how, One of the challenges for me is to catch those projects very early and get them rolling through the process because it has been where we had to resubmit and resubmit and resubmit and resubmit on up to like two months before the project is due, and that was a nightmare. The need to factor in time to respond to IRB feedback is particularly challenging when instructors expect students to conceptualize and implement their projects within one semester. Ideally social work programs offer a two-semester sequence where students submit their IRB proposal during the first semester therefore allowing adequate time to attend to IRB feedback and complete their projects within the second semester. The process becomes even more intensive when a student project must undergo multiple reviews, for example an agency-based review process as well as the university IRB. The potential exists that the reviewing entities will have different or even conflictual requirements. One interviewee described the experiences of a student who went through her university IRB and organization’s review process: ‘‘she had to prove to [her organization that] it was gonna benefit their organization directly…and she did build a case for that but that’s certainly not something that is a standard that the IRB has.’’ As a result of the amount of time required to submit and receive approval, students are at risk of not being able to complete their projects within the given timeframe. A concern was voiced that this ‘‘stunts the learning process.’’ In order to satisfy the IRB requirements in a timely fashion, the potential exists for IRBs to limit student projects to activities that were seen as ‘‘less meaningful’’ or ‘‘superficial.’’ For example, projects that rely upon pre-existing data that prevents students from creating and administering their own surveys, and furthermore may not contain information related to their areas of interest. Lack of IRB Resources A lack of available IRB resources included insufficient access to IRB staff for consultation, as well as inadequate written materials aimed at facilitating the completion of IRB applications. Desired documents included checklists that summarize the requirements as well as guidelines specific to different research approaches. More specifically, one interviewee described how their IRB lacked information regarding how to ethically proceed with Web-based research. Consequently the student had difficulty anticipating IRB requirements, and encountered challenges in gaining IRB approval. While access to IRB staff was a challenge for another interviewee, these ‘‘barriers’’ were reduced with the introduction of an online IRB system. Negative Faculty Perception of the IRB Challenges may also stem from the faculty’s negative perceptions or feelings of intimidation of the IRB, which in turn are relayed to their students. Some instructors may present the IRB as ‘‘bureaucratic in nature,’’ potentially implying that the review process does not contribute to the ethical conduct of research. Others may instill fear in students with stories related to IRB required revisions. These stories may result in students becoming overly nervous or dismissive of the process. The question remains as to whether instructors should share IRB horror stories or gloss over any negative experiences emphasizing only the intended value and purpose of the IRB. Ideally instructors have adequate understanding of and experience with the IRB to communicate a realistic and balanced presentation of the IRB process. IRB Committee Composition and Dynamics A challenge may also surface when the IRB reviewers are not from the social sciences, and particularly when there are no social workers on the committee. The potential exists for the feedback to reflect a lack of understanding of the types of projects developed by social workers. For example, projects that value gathering the perspectives of vulnerable populations rather than relying on provider perspectives may be questioned. One interviewee described the challenges related to committee composition in the following manner, And so you know a chemist might not understand social work research in a way that a social worker does, and so I find that sometimes they’re kind of picky about things that are really not about protecting subjects, but about grammar or something like that. Another expressed concern was the IRB reviewers’ tendency to inflate risk. This is perhaps a result of ‘‘thinking in worse case scenarios,’’ which then requires researchers to identify contingency plans or strategies to prevent/minimize risks that were considered extremely unlikely. Risk inflation may also be a result of IRB reviewers and possibly the review process itself using biomedical research as the standard. This concern was expressed through the use of contrasting statements to indicate the lower risk level of social work studies. For example, one interviewee commented, ‘‘what we do is benign in research compared to, for example, what happens in our medical school.’’ Or as another interviewee shared, ‘‘fortunately in the social and behavioral sciences we don’t have some of the serious physical risks associated with clinical trials, drug trials, medical kinds of research.’’ While some interviewees focused on the lower level of risk as compared to biomedical projects, others acknowledged how social work research poses nonphysical risks especially since studies often recruit vulnerable individuals or groups. As one interviewee shared, In what is considered even benign kinds of research, like satisfaction with mental health services, when we ask people how satisfied they are we ask lots of questions that can be quite anxiety producing or upsetting to folks or provoke some stress. Committee dynamics were also seen as challenging or frustrating when the review process and outcomes appeared inconsistent. Inconsistency manifested itself by similar projects receiving IRB approval at one time with minimal feedback, while at a later point receiving extensive feedback. As a consequence students and faculty have difficulty anticipating IRB concerns. One interviewee described how this is frustrating especially when students structure their IRB proposals using a formerly approved student project as a guide. For some of the interviewees, inconsistency or evolving requirements were caused by IRBs becoming increasingly liability focused and/or overly controlling in their review. At times IRBs were also perceived as overstepping their bounds by monitoring ‘‘research’’ activities in nonresearch courses. For example, when an IRB attempts to monitor practice class assignments that entail interviewing providers. Another cited example of ‘‘overstepping bounds’’ entailed IRBs questioning agency activities or ‘‘policing research methodology.’’ In these instances, a concern was that the IRB drifted from their intended mission, and moved beyond ethical analysis. According to one interviewee, this occurs more frequently with faculty projects while the IRB instead tends to be ‘‘kinder to students.’’ Other interviewees, however, questioned whether the IRB process is geared sufficiently for student projects that often are not funded and smaller in scope than faculty projects. Does this Student Project Require an IRB Review? Prior to determining whether an IRB application is necessary, an iterative process may occur where instructors guide students through a series of revisions to strengthen the proposed methodology and/or minimize the ethical concerns. An interviewee describes the process as guiding the students to consider ‘‘what practitioners [might] think the ethical issues are’’ and how they might revise the proposal to minimize or resolve potential ethical concerns. For example, instructors might encourage their students to conduct anonymous surveys rather than interviews. Students may also be encouraged to provide feedback to their classmates, sometimes in the format of a ‘‘mock IRB.’’ As a means to motivate students to minimize risks, one interviewee shared, ‘‘I’ve usually said these are the hurdles you are gonna have to go through to do this, and that’s often been enough to discourage students.’’ Overall, revisions to a student project may be prompted from input from the research instructor, classmates, or an agency representative (assuming a project is agency-based). Rarely did interviewees identify project ideas that were rejected outright by the instructor prior to IRB submission. An example of a project not allowed entailed interviews with prisoners, as the perceived risks outweighed the potential benefits. Once the instructor feels comfortable with the student proposal, a range of responses emerged regarding the necessity of IRB review requirements. These responses were grouped in the following manner: (a) all projects require a review and (b) all projects meeting specific criteria require a review. As a follow-up question, interviewees were asked to identify projects where they believed no IRB review was required, or were less than certain as to whether a review was required. One interviewee referred to this later category of projects as ‘‘grayer area projects.’’ All Projects Require a Review For some interviewees, every student project must be submitted to the IRB or at the very least a department-level review committee. An independent review can provide a more objective assessment of a project and may identify risks overlooked by the instructor. As one interviewee reflected, ‘‘there has been times in that departmental review board where my colleagues have suggested reasons for the research being sensitive that had not yet occurred to me.’’ Other reasons for an independent review included reference to the regulations’ stance regarding who can determine whether a project is actually exempt. As one interviewee shared, The decision as to whether something is exempt or not needs to be with an entity that is very clear in terms of what the rules are and it shouldn’t be left up to individual’s own discretion. I mean I think its fine for someone to argue that their project is exempt, but I think someone else should make that determination. A concern was voiced that submission of all student projects may result in an IRB work overload. Perhaps if this occurs, social work programs and their IRBs would be motivated to negotiate a policy regarding requirements for student projects. All Projects Meeting Specific Criteria Require a Review For other interviewees a project should be submitted to the IRB only when it meets specific criteria. For these interviewees, general consensus emerged that if the response were ‘‘yes’’ to the following questions, an IRB review would occur: Does the project meet the human subjects regulations’ definition of research? Is it more than minimal risk? Does it involve vulnerable populations as defined by the regulations (e.g., prisoners, children)? No Review Required or ‘‘Grayer Area Projects’’ Some interviewees were unable to identify ‘‘grayer area projects.’’ Three main reasons for the clarity regarding IRB review requirements emerged: (a) university or department guidelines existed that clearly specified review requirements, (b) belief that all projects must be reviewed, and (c) clear perception regarding types of projects that fall outside the scope of the review process. The types of projects that were perceived by some interviewees as falling outside the IRB’s purview included classroom projects completed for educational purposes only, program evaluations, and opinion surveys of a ‘‘benign nature.’’ As an example, one interviewee described not requiring IRB approval for projects ‘‘where students will conduct a survey on each other that does not require divulging personal or sensitive information.’’ Other interviewees, however, identified ‘‘grayer area projects.’’ When confronted with ‘‘grayer area projects,’’ some interviewees consulted with their colleagues and/or IRB, while others automatically required their students to submit an IRB application. For example, one interviewee reported being unclear as to whether projects using Census Data required IRB approval. Consequently he required that his student submit an IRB application. For some interviewees, program evaluations represent one form of ‘‘grayer area projects.’’ Interestingly a wide range of responses emerged regarding whether program evaluations required IRB review. On one end of the continuum, some interviewees would submit program evaluations given their stance that the IRB determines whether a project is actually exempt. As one interviewee shared, the possibility also exists that a researcher might have an overly broad conceptualization of program evaluation in order to avoid IRB submission. On the opposite end of the continuum, others argued that evaluation projects do not require IRB approval based upon the regulation’s definition of research that emphasizes generalizable knowledge. Hovering between these two positions were other interviewees who were uncertain of the requirements. Other examples of ‘‘grayer area projects’’ included: agency led versus student led projects (e.g., the agency identified the question), agency-based research that utilizes de-identified secondary data only, anonymous surveys to providers, and projects where students assist a faculty member on a study that already has IRB approval. Perhaps in anticipation of research instructors not wanting to submit their student projects through the IRB review, one interviewee pointed out that ‘‘grayer area projects’’ are likely to be lower risk studies that do not necessarily require a full committee review. Recommendations to Facilitate the Review of Student Projects Interviewees were asked for suggestions on how to facilitate the review of student projects. Three basic recommendations emerged: policy development, research instructor training in human subjects protection, and suggestions for IRB reviewers. Policy Development The policy or guideline recommendations consisted of suggestions for checklists that summarize the required information to complete an IRB application, as well as information that outlines what is not required (e.g., an extensive review of the literature). Additionally a recommendation emerged for the development of a protocol specifying the types of student projects requiring a review. A suggested format for the protocol included a decisionmaking tree that would indicate whether a review was required and/or what level of review might be needed. For example, a series of questions might be: Is the student analyzing pre-existing data? If yes, is the data de-identified? If yes, are the findings intended to remain within the agency and/or the student’s classroom? If yes, then some institutions may require an exempt review. Decision-trees must be approved by the IRB. The policy could also reiterate that the research instructor reviews for ethical and methodological soundness of all student projects regardless of whether they must go through the IRB process. Suggested areas for research instructors to focus upon in their reviews included risk/benefit analysis, consideration of cultural appropriateness, and absence of coercion. While a protocol might minimize the ambiguity around student review requirements, a concern was raised that the rigidity of a protocol might impede an instructor’s ability to exercise some discretion regarding review requirements. Beyond creating a decision-making tree, others pointed to the value of negotiating directly with the IRB to determine what types of student projects can be automatically waived from an IRB review. Suggested examples of projects that could be waived included strict program evaluations and projects aimed at assessing or improving the curriculum. Alternatively, course waivers might be secured. In these instances instructors would demonstrate how they completed the appropriate training to ‘‘insure that these projects follow university regulations as far as protecting human subjects.’’ With a course waiver, the instructor would not submit the individual student projects to the IRB, and instead would take full responsibility to assure ethical conduct and that the projects remain at an appropriately low risk level. Required Research Instructor Training Instructors should complete an online ethics certification program demonstrating their knowledge regarding the basic ethical principles, the criteria specific to the different levels of review, and the overall human subjects regulations. By having a fuller understanding of the exempt criteria, for example, instructors could then guide students in the design of exempt-level projects. This would allow quicker IRB reviews and more importantly would help assure that the student projects posed very minimal risk, which some interviewees argue is appropriate given students are just learning how to do research. Suggestions for the IRB A suggestion emerged for IRBs to become more familiar with the course structure surrounding student projects, especially those that occur in the context of field placements. This would allow IRBs to understand how students receive multiple layers of supervision including from their field and research instructors. Increased social work representation on IRB committees would enhance IRBs’ understanding of social work projects. A second recommendation was for IRBs to devise a system that would allow for a timelier review of student projects. Ideally the creation of this system would be a result of a collaborative effort between schools of social work and their IRBs. One suggested approach was for a departmental or IRB sub-committee to review student projects. If a project raised red flags (e.g., in terms of risk), then a more time-consuming full review would be required. A final recommendation for the IRB was to develop a mechanism to alert researchers in a timely manner to regulatory changes. Awareness of changes can assure that instructors guide their students in a more accurate manner ideally allowing for better compliance with the IRB requirements. Other recommendations entailed improving coordination with other IRBs (when a project requires multiple-IRB reviews) and increased IRB access for students. DISCUSSION The interviewees identified both benefits and challenges with the IRB process, as well as specific recommendations to facilitate the review of student projects. Overall the benefits contribute to students becoming ‘‘socialized’’ into taking research ethics seriously and perhaps shifting from a tendency to conduct research haphazardly. The primary recurring challenge pertained to issues of time. The time-consuming nature of the IRB is discussed within the literature in the context of student projects (e.g., Foss, 2005) and for researchers in general (e.g., Shore & West, 2005). Similar to the recommendation voiced by Foss, some interviewees suggested students complete the IRB process in their first research course, thus allowing adequate time to gain approval and implement their projects during their second research course. Mutual education and negotiated efforts capture an overarching thrust of the different interviewee-generated recommendations. Research instructors are encouraged to gain greater understanding of the human subjects regulations and IRB process, while the IRB is encouraged to familiarize themselves with the types of projects social work students often conduct. Through this mutual education, the IRB and social work faculty are better positioned to negotiate a method for the review of student projects. Interviewees who reported no challenges with their IRBs at times alluded to processes already in place that facilitate the review of student projects, such as course waivers or streamlined review procedures. Unfortunately the interview questions did not explore whether these structures were negotiated or determined by the IRB. Regardless, models exist that social work programs could explore with their IRBs. A departmental review board process is one intriguing approach referenced by several interviewees. Departmental level review boards comprised of social work staff and faculty may increase the likelihood that the reviewers consider a fuller range of ethical considerations by integrating the NASW Code of Ethics (1999) with the traditional IRB considerations. Additionally these reviewers likely will have a greater understanding of the types of projects that students conduct, particularly those that occur within the field placement. These factors could minimize the expressed challenge that University IRBs do not always provide feedback that demonstrates an understanding of the proposed student projects. Having reviewers housed within the department may also increase accessibility for students and research faculty allowing for more expedient clarification of areas of confusion. This may ultimately result in a more timely review process. Departmental review committees also could recruit advance year social work students, thus providing an ongoing opportunity for students to exercise their knowledge and skills related to ethical research. Requirements for student reviewers could include having submitted their own proposal for an ethics review and completion of an online human subjects certification program. If the departmental review board identifies ethical concerns with a project, they can offer the student and instructor a set of options. These options may entail having students revise their projects to minimize the risk or forwarding the proposal to the university-wide IRB. Projects that pose greater than minimal risk should automatically be directed to the IRB. A challenge to departmental reviews includes the possibility that social work programs have inadequate resources including faculty/staff expertise and time. Schools might also encounter resistance from members of the social work faculty as well as their IRBs. Some faculty may question whether a departmental review dilutes the purpose and value of the IRB, producing less rigorous assessments of student projects. Other faculty may question the need for a departmental review board if they perceive the IRB as functioning effectively. Several interviewees reported experiencing no challenges with their IRBs. If a departmental review process is desired, schools must collaborate with their IRBs to determine feasibility and to agree upon types of projects that could be reviewed at the departmental level. These agreements must factor in the human subjects regulatory requirements and Office for Human Research Protections’ guidelines, including how the level of review for research proposals is determined and who can determine whether a project is exempt. If it is not feasible to create a departmental review board, perhaps schools of social work can work with their IRBs to create a more streamlined process for the review of student projects that pose no greater than minimal risk. As a possibility, the IRB could have a designated sub-committee or staff person for the support and review of student projects. Schools must work with their IRB if there is an interest in developing a model for the review of student projects. The outcome of these negotiations will likely differ, as IRB requirements may vary across institutions (Jenson et al., 2003). Differences may emerge depending upon the volume of research produced at a given University. Smaller university and college IRBs presumably will have less IRB infrastructure and perhaps less specified IRB review protocols. Based upon the study findings, some research instructors may embrace increased specification of review requirements as this would eliminate or at least decrease areas of confusion. For example, some research instructors may welcome greater specification regarding IRB requirements of program evaluation projects. Others, however, may view increased specification as leading to a decrease in instructor discretion to assess review requirements. FUTURE STUDY DIRECTIONS The findings from this study are based upon the experiences and perspectives of 16 graduate research faculty. By expanding the sample the possibility exists that other IRB challenges, benefits and recommendations will be identified. The findings from this study, however, coincide with the sparse literature related to student projects and the IRB. Another future study direction includes soliciting the perspectives of field instructors, given that many of the student projects occur within their field placements. The field instructors can offer insight regarding the overall benefits and challenges in having students conduct agency-based evaluations. Specific to the IRB process, field instructors could identify what kinds of resources would assist them in guiding students in their research endeavors. Foss (2005) described online resources provided to public health nursing field instructors, which included information on the IRB process as well as a human subjects regulations tutorial. Questions for field instructors should also focus upon whether their agencies have an internal review board, and if so how this process compares or complements the university review board. One of the identified challenges was the difficulty in navigating multiple review boards. A content analysis of different university IRB protocols for the assessment of student projects represents yet another possible future study direction. Questions could include how many universities have a studentspecific protocol, what are the similarities across protocols, and where are the divergences? Descriptive variables related to the different universities should also be collected (e.g., size of university, volume of research produced, degree of IRB infrastructure). Key informant interviews (e.g., IRB administrators) could supplement the analysis of student protocols allowing for greater insight regarding how the protocols were developed, who was involved in the development, and what are the strengths and challenges of the current protocol. For those universities that do not have an established protocol, interviews could explore why a protocol does not exist, and whether there is a perceived need for a protocol. The issue of risk represents an additional area for future inquiry. Across numerous interviews, comments were made regarding the perceived lower risk of social work studies as compared to biomedical research. While there may be less physical risks in social work research, one should not assume that social work research is risk-free. Boothroyd and Best (2003) contend that we need a more critical understanding of the risks involved in social work research. These authors note a dearth of empirical studies that critically examine the probability and severity of risk associated with social work studies, including a lack of information based upon actual study participants’ perceptions. They argue, as did some of the interviewees from this study, that social workers need to consider possible risks to study participants who often represent vulnerable individuals or groups. We must balance the value of including often-marginalized voices in research, with the need to create a respectful and ethical process. Different directions exist to build off of this exploratory study. Social work programs are encouraged to engage in dialogue with their IRB to explore options to facilitate the review of student projects. Negotiation efforts should not lose sight of the shared goal of promoting and assuring ethical research.

An intensified quest for so-called excellence in public research systems in the Western world over the past couple of decades has shown itself palpably in the introduction of new, targeted, funding programs by governmental and private sponsors alike (Braun 2003; Cozzens 2007; Mu¨ nch 2007; Wildavsky 2010; Whitley 2011). These excellence funding programs are typically aimed at singling out the most worthy and promising recipients of funding, be they individuals, groups, research environments, or organizations, in order to support the most competitive research activities and ensure the long-term quality of national R&D efforts and, by extension, secure future economic growth. In Sweden, the rather one-sided promotion of Centers of Excellence of the past decade, that is, increased support to specifically chosen large research environments (Hellstro¨m 2011, 2012; Hallonsten and Silander 2012) appears to have been complemented by a focus on excellent individuals— ‘future research leaders’ in the vocabulary of one sponsor— in governmental policy as well as on the discretionary level of public and private funding bodies. The shift is practical but also broadly discursive: Several new funding programs for excellent young (postdoc level) researchers have been launched, including a governmental program detailed in the most recent governmental research bill, and all the calls seem to build on the assumption that there is a structural deficit in the Swedish public research funding system that leaves early career researchers without proper funding and career advancement opportunities. The performer side of the Swedish public research system is dominated by a large academic sector, encompassing the old universities as well as newer and smaller (regional) colleges. On the supply side, the system is decentralized and loosely coordinated, especially the system for third-party funding which has grown significantly in the past few decades and been transformed through two recent major policy reforms (Benner and Sandstro¨m 2000b; Engwall and Nybom 2007; Benner 2008; Hallonsten and Holmberg 2013). The lack of coordination has created inefficiencies as well as by-default resource concentration in the older universities, and has created a cluttered and opaque policy field. Both the education and research missions of the academic sector have been expanded, and simultaneously, academic governance and career structures have been deregulated, including the abolishment of the chair system and the subjugation of the professoriate and all other faculty to general labor law and the central management and budgetary work of the universities (Engwall and Nybom 2007; Hallonsten and Holmberg 2013). For individual academic teachers and researchers, these developments have partly countered the concurrent increase in governmental funding for research in the universities and created a situation of perceived harsh competition for research funding and academic positions. In the expanded organizational field of research funding, many new organizations have shown ambivalence, lack of focus, and built-in conflicts of interests, as well as a certain inclination to very swift incorporation of new missions and policies as these emerge on political level or as initiatives among the older and more reputable funding agencies (Sandstro¨m 2000; Benner 2003; So¨ rlin 2003; Hellstro¨m and Jacob 2005). In this article, we analyze the pluralism and lack of coordination in the Swedish research funding system and its effects on the level of a specific type of funding programs that have been given some attention recently, namely the support of those young academics that are judged to be especially promising, thus worthy of generous support, and also somehow structurally discriminated against. The argument has two parts and the analysis two corresponding levels. First, we show that there is striking uniformity in aims and purposes of these funding programs, and we use the concept of institutional isomorphism from neoinstitutional theory (DiMaggio and Powell 1983) to explain this uniformity, which appears quite surprising at first sight, given the apparent pluralism and diversification of the field of funding organizations and the relative youth of several of these organizations. Second, we use data on some recent rounds of these funding programs and the recipients of grants, to show one consequence of such uniformity, namely agglomeration effects, i.e. the attraction of more than one of these grants by the same individual. Such effects are clearly unintended: The stated purpose of all these programs is to mitigate a structural deficit in the system that allegedly leaves young (postdoc level) researchers without funding at a crucial and vulnerable career stage, and each program is designed to provide the full support for several years for a promising individual to build up a research base. Concurrently, and this is where the two parts of the argument are connected, these inadvertent effects are expectable, given that research funding is inherently selective and competitive and coordination between different funding bodies typically weak. Institutional isomorphism explains, quite forcefully, the logical step between such lack of coordination in a field and a certain uniformity of the actions on behalf of the organizations in the field, and the article thus also contributes by importing a theoretical concept from organizational sociology into studies of research funding and research evaluation, and proving its explanatory value. Importantly, the intentions are not normative: We do not use the demonstrated agglomeration of funding as an argument for painting the Swedish public research funding system as inefficient, bungled, or discriminatory on structural or individual level. Our intentions are descriptive and conceptual, i.e. to display and discuss the effects of weak coordination and discursive homogeneity in pluralist research funding system, with the aid of the operationalization of a forceful theoretical model. 2. Background and theory Sweden, as most Western countries, expanded its public research system significantly in the few decades following World War II, and established a mix of first-stream and second-stream funding programs through the institutionalization of governmental block grant funding and the creation of several disciplinary research councils. On the performer side, the system remained heavily dominated by the universities also through the expansion of higher education, compared with many other countries where strong research institute sectors were built up to complement academia. The universities and colleges in Sweden together account for approximately 75% of the governmental research budget, with the remainder shared by industrial R&D (approximately 15%) and nonacademic public research (below 10%) (Statistics Sweden 2013). In spite of wide-ranging restructuring of academia in the past few decades, including managerialism reforms, abolishment of the chair system, and changes to the academic governance structures, classic academic norms and ideals remain strong. This is mainly due to the relative domination of the universities founded in the 19th century and before, who outperform the more recently established regional colleges both in resource strength and research productivity (Hallonsten and Holmberg 2013). Though never an explicit policy objective (quite the reverse), this stratification seems only to have increased over time, due to a combination of increasing competition in the funding system, the recent profound policy focus on excellence, and heavy 250 . O. Hallonsten and O. Hugander organizational path dependence (Hallonsten and Silander 2012; Hallonsten and Holmberg 2013). The funding system has undergone substantial restructuring and diversification in the past two decades. In addition to the general relative growth in competitive funding for research in academia, common to most (Western) countries, two major policy reforms are important. First, the early to mid 1990s founding of a number of new public research foundations out of the so-called ‘wage earners funds’,1 all with specific mandates such as internationalization, academy–industry interaction, and support for research in the regional colleges, all missions quite typical for the time (Benner 2005). Second, the 2001 restructuring of the research councils by which the model of four disciplinary research councils2 and a National Council for Planning and Coordination of Research was abandoned in favor of a mission-oriented model of one larger, broad-encompassing Swedish Research Council (Vetenskapsra˚det, VR), a separate innovation agency (the Swedish Agency for Innovation Systems, Vinnova hereafter), and two research councils in areas considered of special priority for Sweden (Engwall and Nybom 2007: 42).3 Today, the total annual (2012) amount of approximately 33 billion SEK (corresponding to approximately E4 billion) of research income in the Swedish universities and colleges comes from a rather broad variety of sources, of which the governmental base grant is the largest with 39.3% of the total research income, but with the rest—i.e. more than half—divided on a number of grant-allocating bodies and contractors of research. Competitive third-party funding from public sources within Sweden—the research councils, other governmental agencies (including Vinnova), and the public research foundations—amounts to approximately 20%. The 10 largest of these public funding bodies allocating grant money on open and competitive basis are listed in Table 1 along with information on their founding years and organizational history.4 This rather pluralist and twice newly restructured field of public research funding organizations shows strong signs of what has been canonized within neoinstitutional theory as the force of institutional isomorphism in organizational fields (DiMaggio and Powell 1983). Institutional isomorphism is a theoretical concept for explaining why groups of organizations operating in a common area (e.g. business area or societal sector) and with similar general practices, become gradually more alike over time despite preconditions (e.g. specialization, competition) that would seem to be a driver of an opposite development toward differentiation. The desire of organizations to establish legitimacy for their activities (also relative to competitor organizations in the field), is typically what drives isomorphic change, since especially new organizations obtain legitimacy largely by adopting forms and practices of older, established organizations in the same field. Legitimacy is field-specific and has to do with legislation, norm systems, traditions, and established practices in fields (Meyer and Rowan 1977: 352; Suchman 1995: 574). The concept of institutional isomorphism has been used by several scholars to explain systemic shifts in public science systems under change and growth, most of all the behavior of new universities and higher education institutions in search of legitimacy (Maasen and Potman 1990; Sko¨ ldberg 1991; Stensaker and Dahl Norga˚ rd 2001; Morphew 2002; Morphew and Huisman 2002; Rusch and Wilbur 2007), but also older or well-established universities seeking to cope with broad policy changes and the global restructuring of higher education systems (Halffman and Leydesdorff 2010; Wilkins and Huisman 2012; Frølich et al. 2013). The application of this theoretical concept on other parts of public science systems, for example, as a tool to explain changing behavior of research funding bodies and other research policy actors has, however, been rather sparse. Those studies that exist point out that organizational actors in public science systems are very susceptible to institutional isomorphism because of the strong norm systems of science which are institutionalized in academic organizations and sustained by classic governance systems of science such as peer review and the execution of academic freedom in teaching and training of doctoral students, etc. (Halffman 2005; Luukkonen 2014: 32). For fields of research funding organizations, which to our knowledge have not been studied specifically within the framework of neoinstitutional theory, institutional isomorphism would mean that funding bodies seek legitimacy by launching funding programs that support a particular kind of excellence that is promoted in current policy discourse, and seek to imitate what other (older) funding bodies are doing. Furthermore, especially in small countries, several funding bodies will most likely have to draw upon the same relatively small group of peer reviewers, which means that they will incorporate institutionalized evaluative standards. In other words, we may expect that a broad consensus of research excellence will run across many funders and programs. In Sweden, such tendencies have indeed been observed, and the case is especially interesting since several of the large funding bodies (see Table 1) are public and private foundations and thus independent from governmental policymaking and, furthermore, were once founded with specific missions that deviate from the mainstream research funding mission of e.g. the research council(s) (Hellstro¨m and Jacob 2005; Benner and So¨ rlin 2007; Hallonsten and Silander 2012). The concept of institutional isomorphism offers a model to explain this behavior of seemingly independent organizations that are both self-governing by regulation and quite specialized in missions, to adopt the forms and practices of each other. The original conceptualization by DiMaggio and Powell (1983) specified three types of isomorphism that are partly overlapping and somewhat difficult to separate in empirical analysis: coercion, normative pressure, and mimesis. In the case of research funding systems, coercive isomorphism is expected to be primarily driven by governmental legislation and oversight; normative pressure is the adherence by new actors to norm systems that are vivid in a system; and mimesis is a response of (new) funding bodies to a situation of uncertainty with regard to long- and short-term missions and purposes of their activities. Governmental funding bodies (e.g. research councils) are quite naturally exposed to coercive isomorphism due to legislation and governmental oversight. Moreover, it has been shown that funders typically establish legitimacy around new funding programs by making them sustain current norms of research systems (Benner and Sandstro¨m 2000a; Benner and So¨ rlin 2007). Finally, the altered societal demands and expectations on science as well as the current institutional restructuring and new global competition in public science systems, has created uncertainty with respect to missions, goals, and quality standards for academic science, which affects not only the performers of research but also funders (Hessels et al. 2009; Jacob 2009; Hellstro¨m 2011; Stephan 2012). Funding bodies can be expected to respond to this uncertainty by seeking to establish legitimacy by mimesis. The Swedish research policy and funding system, diversified and lacking much of the coordination seen in other comparable countries (Benner and Sandstro¨m 2000b; Engwall and Nybom 2007; Benner 2008; Hallonsten 2011; Hallonsten and Holmberg 2013) can be expected to be especially susceptible to institutional isomorphism, in spite of seeming diversification and functional specialization. 3. Supporting ‘future research leaders’ in Sweden The most recent trend in Swedish research funding to promote especially excellent young individuals has emerged out of a seemingly unanimous policy analysis among both public and private funders that the Swedish public research system is in severe lack of proper career enhancement opportunities for scientists on postdoc level, including unsatisfactory competitiveness in the process of obtaining third-party funding (e.g. Barkeman 2010; SSF 2010; Nilsson 2012). The most recent governmental research bill (2012), a quadrennial document that both lays out the governmental policy objectives and summarizes the policy analyses of the several stakeholders that are invited to give their input in advance of the writing of the bill, identifies a threatening structural deficit in the public research system and its funding landscape, in need of correction: ‘The careers of young researchers are often inhibited by shortterm and inadequate funding of research projects. These young researchers thus have a hard time liberating themselves from previous supervisors and established groups in order to build an independent research line. There is, hence, a need for sufficiently large research grants for younger researchers so that they can be given the opportunity to independent production of results that can give them the qualifications needed for obtaining permanent positions at Swedish higher education institutions and thus enable further careers as researchers.’ (Government bill 2012: 67–8). 252 . O. Hallonsten and O. Hugander The funding programs under study here have all been launched on basis of similar analyses, i.e. that promising young researchers lack opportunities to obtain funding necessary at a critical career stage, and the programs are therefore all designed to mitigate this problem. The programs under study include three Swedish funding programs as well as one European (Table 2). Included in the analysis is also regular funding from the Swedish Research Council in order to (partially) try the assumption that the recipients of these grants are discriminated against in the funding system. The Future Research Leaders program of the Swedish Foundation for Strategic Research (SSF hereafter) was launched in 2000 as a renewed piece in the execution of the foundation’s mission to ‘promote the development of strong research environments of highest international standard with importance for the development of Sweden’s future competitiveness’ (SSF 2001: 2). The aim of the program is to ‘support and promote young scientists who have the potential and the ambition to become future leaders of academic and/or industrial research in Sweden’, and it is open to Swedes and foreigners alike and supports a stay at a ‘Swedish research organization—university or public or private non-profit research institute—of the choice of the grantee’. The funding is intended to allow the grantee to build up a research group, and so only a maximum 50% of the grant can be used to cover the grantee’s own salary (SSF 2009). Calls are issued every 3 years; the two calls included here are those of 2008 and 2011. The SSF is also the organization behind the Ingvar Carlsson Award, named after the former Swedish prime minister who was also the chairman of the foundation between 1997 and 2002. The award was launched in 2005 after the foundation (way ahead of the government and the other funders appearing in the material) had identified a ‘bottleneck’ in that early phase of scientific careers of supposed ‘transition from postdoctoral studies to independent research’. The aim of the program is ‘to identify and support young, well-qualified postdocs who intend to start independent, lasting and creative research careers on their return to Sweden’, and the proposed research ‘should have a potential to strengthen Sweden’s future competitiveness’. Grantees must have a PhD from a Swedish university (not older than 4 years) and spent at least 12 months as a postdoc abroad, and the grant is upon condition that the grantee becomes ‘permanently active at a university in Sweden’ (SSF 2010). Calls are issued every 3 years; included here are the calls of 2009 and 2012. The ERC Starting Grants were launched by the European Research Council (ERC hereafter) as its first major program upon the council’s founding in 2007 on basis of the analysis that the available funding opportunities in Europe are quite limited for ‘researchers who have recently (2–9 years ago) completed their doctoral studies to make the transition to independent research leaders in their own right’, a ‘crucial stage’ in their careers (ERC 2007). Consequently, the Starting Grants ‘aim to support up-and-coming research leaders who are about to establish a proper research team and to start conducting independent research in Europe’. Calls are annual and there are no restrictions on, or specific quotas for, scientific disciplines of the applicants. The research must be carried out in a public or private research organization within the EU or an EU Associated Country (ERC 2013a). For the five calls in the history of the grant (annually 2007–12), a total of 23,364 applications have been received, with a record high turnout in 2007 (9,167 applications) and an average success rate of 10.7% (ERC 2013b). Although the analysis of this article is focused on Sweden and the data exclusively Swedish, the inclusion of the ERC Starting Grants in the analysis is conceptually warranted: In the current geopolitical context, the pan- European funding body ERC can rightfully be identified as a role model organization (or at least a very influential one) for Swedish national funding bodies. Importantly, this does not at all mean that the argument about institutional isomorphism is thereby extended to Europe but rather that European policymaking and the paragon Table 2. The funding programs studied Funding program Funding body Rounds included Grant amounts and time periods SSF Future research leaders Swedish Foundation for Strategic Research (SSF) 2008, 2011 SEK 8,5–10 million (& E1–1.2 million) over 5 years Wallenberg Academy Fellows Knut and Alice Wallenberg Foundation (KAW) 2012 SEK 5–7,5 million SEK (& E0.6–0.9 million) over 5 years. Ingvar Carlsson Award Swedish Foundation for Strategic Research (SSF) 2009, 2012 SEK 3 million (& E0.36 million) over three to 4 years. ERC Starting Grants European Research Council (ERC) Annual calls 2009–12a Averagely E1.5 million (& SEK 12.5 million) over up to 5 years. Regular Research Council funding: project grants, employment grants, framework grants Swedish Research Council (VR) Annual calls 2008–12 Project grants average (2012) SEK 3.6 million (& E0.43 million) aThe ERC Starting Grants are issued annually since 2007 but for reasons unexplained, the online database of grantees does not cover the 2008 call. example of European research funding trends as embodied by the ERC (quite duly) is included as one important source of influence over Swedish research funders. The Wallenberg Academy Fellows program is the most recent of the programs, launched by the Knut and Alice Wallenberg Foundation (a private foundation; KAW hereafter) in 2012. Advertised as a ‘career program for Sweden’s most promising young researchers’ aimed to give them ‘long term resources so that they can attach difficult and long-term research questions’, the program is the foundations largest ever with an expected total amount of SEK 1.2 billion (approximately E140 million) to be allocated during a total of 10 years (KAW 2013a). Successful grantees must be awarded a permanent position as assistant professor at the host university, and the universities must guarantee that at least 50% of the researcher’s payroll expenses be covered by internal grant co-funding (KAW 2013b). Finally, the Swedish Research Council’s annual calls for project grants, employment grants, and framework grants vary somewhat from year to year, although the regular call for grants is almost identical in scope and size from one year to another. As seen in Table 1 above, the Swedish Research Council is the largest third-party funder of research in Swedish academia and normally allocates funding broadly and solely on basis of scientific quality, although specific programs to young researchers or in specific areas also exist. Included in the analysis here are regular project grants, which the Swedish Research Council awards around 1,000 annually (of a total value of E250–450 million), as well as grants issued through special calls (including framework grants), and previous funding programs such as grants for postdoctoral employment and similar, all for the years 2008–12. Funding from the Swedish Research Council’s general annual call is of course not support directed specifically to young and promising individuals, and it might therefore seem like an inconsistency to account for funding from the Research Council’s general programs allocated to the 130 individuals in the sample. However, since the funding is extremely competitive5 and regarded as perhaps the most prestigious research grants in Sweden, any presence of such grants in the funding portfolios of the individuals in the sample is a good litmus test on whether these 130 individuals actually fit the description of young researchers whose careers are ‘inhibited by short-term and inadequate funding’ and who consequently ‘have a hard time liberating themselves from previous supervisors and established groups in order to build an independent research line’ (Government bill 2012: 67–8, op. cit.). Furthermore, we use the data on council grants selectively, to show that several of the recipients of ‘future research leaders’ grants had obtained highly competitive (and quite generous) council funding before they were chosen to be receiving the ‘future research leaders’ grants in question. Before we proceed to analyzing the data, two additional remarks on the material are necessary. Firstly, the time frame has been chosen to the past 5 years (2008–12) in order to make the material coherent and manageable; thus, for programs that have been in place longer, we have only analyzed recipients of grants in calls from 2008 and on. Secondly, the 130 individuals who have been identified as recipients of any of the four programs in Table 3 (i.e. not counting the Swedish Research Council) may well have other additional funding that we are not aware of, since it is an unfeasible task to make a complete and comprehensive review of all the funding attracted by one individual, let alone 130.6 The analysis of the data proceeds as follows. First, for reasons of clarity and simplicity, we focus specifically on the two largest (in average grant size) Swedish programs, the SSF Future Research Leaders and the Wallenberg Academy Fellows programs, and we analyze how recipients of these two key programs fare in other competition for funding, in order to get an indication of the extent to which these individuals are actually severely underfunded, as the rhetoric underpinning the programs suggests. We then move on to make a broader analysis of the full set of 130 grantees of all the programs in Table 2 (except council grants) in all the years covered, and we then put focus on the group of 22 extreme cases who have received two (or more) different grants among those studied (not including council grants), in order to take the analysis one step further and prove that in some cases, the effect of the programs run squarely counter to their stated purposes and seem to equip individuals with enough resources for a whole excellence center or similar. Finally, we add to the mix a time variable, by which we are able to single out from the original sample of 130 no less than 36 individuals (only with 14 overlaps with the extreme cases) that have received grants within the programs and rounds studied while already in preceding years having secured either another of the grant types studied or general Swedish Research Council funding. Table 3. Figures on recipients of Wallenberg Academy Fellows grants Total number of recipients 30 Number of recipients who also got: SSF future research leaders (2008, 2011) 4 (13.3%) Ingvar Carlsson Award grant (2009, 2012) 1 (3.3%) ERC Starting Grant (2009–12) 8 (26.7%) Any of the above 11 (36.7%) More than one of the above 2 (6.6%) Swedish Research Council funding (2008–12) 21 (70.0%) Any of the above (including Swedish Research Council) 23 (76.7%) 254 . O. Hallonsten and O. Hugander 4. Inadvertent funding agglomeration Using publically available data about grantees, participants and fellows of the Wallenberg Academy Fellow program and the SSF Future Research Leaders program, we now proceed to analyzing practical effects of these funding schemes. Tables 3 and 4 show the results of this analysis. While the agglomeration effects are not as emphasized when considering the overlap of any two programs (the highest overlap rate, 34.2%, being the share of recipients of SSF research leaders grants who also got an ERC starting grant), they are quite obvious when joining the figures together and introducing a time variable. Table 5 shows the same data as Tables 3 and 4, but stripped of some information, so that causal relationships surface. There we see that no less than two thirds of the recipients of Wallenberg Academy Fellows grants in 2012, as well as two thirds of the recipients of SSF future research leaders in 2011, managed to obtain regular Swedish Research Council funding in the preceding 4 and 3 years, respectively. The agglomeration of funding at this level is not particularly striking, although the success of a number of already well-funded (through regular Swedish Research Council grants) individuals in getting the Wallenberg Academy Fellows and SSF Future Research Leaders grants is indeed remarkable. Turning now to the extreme cases, we have identified 22 individuals (out of the total sample of 130) who have been the recipients of two or more of the grants in Table 2 (i.e. not including general Swedish Research Council grants). These make up less than one fifth of the total sample but emerge as particular cases in point. As seen in Table 6, which lists these 22 individuals (anonymized and designated with letters A to V for identification), the most successful among them have collected individual research grants of several million Euros in total. It shall be mentioned that the analysis here naturally does not include other funding—not least internal university funding—which is expectable in almost every case, not only among the 22 extreme cases in Table 6 but in the whole sample of 130 individuals. Both the SSF Future Research Leaders grant and the Wallenberg Academy Fellows grant require the host institution to provide or arrange for at least 50% of the salary for the grantee which means that, for example, person B in Table 6 will have to have an arrangement with her/his host institution that guarantees an additional sum in the region of 1.5–2 million Euros that is earmarked to this individual, leaving her/him with a total sum of research funding amounting to more than five million Euros.7 And, as noted, we do not know other possible additional funding streams going to the 22 researchers in Table 6. It is thus quite clear that some individuals, although they make up a small fraction of the overall data sample analyzed here, are enormously successful in attracting grants, and would typically therefore not qualify as ‘inhibited by short-term and inadequate funding’ and/or as having ‘a hard time liberating themselves from previous supervisors and established groups in order to build an independent research line’ (Government bill 2012: 67–8, op. cit.). But surely, the connection between these displayed agglomeration effects and the lack of coordination in the organizational field of Swedish funding bodies is not sufficiently established just by the fact that some individuals come out as extraordinarily successful. Such phenomena can, conceivably, be attributed partly to luck and not least perhaps to sudden career ‘jumps’ that make individuals succeed at several fronts simultaneously, and funders cannot be expected to coordinate ongoing evaluation and selection procedures. However, when adding a time variable, the argument becomes more convincing. Table 7 shows 36 individuals in the sample8 who all received grants within the programs and rounds studied while already in preceding years having secured either another of the grant types studied or general Swedish Research Council funding. The overlaps with the extreme cases in Table 6 are only partial, i.e. only 14 individuals at the top of Table 7, who bear the same capitalized ID letters that identify them in Table 6. Some individuals emerge again as extremely successful; e.g. individual H who Table 5. Selected figures with added time variable Number of recipients of the 2012 Wallenberg Academy Fellows grant (total 30) who also got: SSF future research leaders (2008, 2011) 4 (13.3%) Ingvar Carlsson Award grant (2009) 1 (3.3%) ERC Starting Grant (2009–11) 4 (13.3%) Swedish Research Council funding (2008–11) 20 (66.7%) Number of recipients of the 2011 SSF Future research leaders grant (total 18) who also got: ERC Starting Grant (2009–10) 4 (34.2%) Swedish Research Council funding (2008–10) 12 (66.7%) Table 4. Figures on recipients of SSF future research leaders grants Total number of recipients 38 Number of recipients who also got: Wallenberg Academy Fellows grant 4 (10.5%) Ingvar Carlsson Award grant (2009, 2012) 0 (0.0%) ERC Starting Grant (2009–12) 13 (34.2%) Any of the above 18 (47.4%) More than one of the above 4 (10.5%) Swedish Research Council funding (2008–2) 34 (89.5%) Any of the above (including Swedish Research Council) 34 (89.5%) Supporting ‘future research leaders’ in Sweden . 255 managed to obtain a Wallenberg Academy Fellow grant in 2012 after already having secured council grants in 2009 and 2011, and an ERC Starting Grant in 2010; or individual L who has a similar history with a council grant in 2008 and an ERC Starting Grant in 2010. The perhaps most striking result shown in Table 7 is that no less than 70% (21 of 30) of the recipients of the most recently launched program under study here, the Wallenberg Academy Fellows, have previously obtained either an SSF Future Research Leaders grant, an ERC Starting Grant, general Swedish Research Council funding, or an Ingvar Carlsson Award. For the 2011 recipients of the SSF Future Research Leaders grants, the same figure is two thirds (12 of 18). Obviously, the number of extreme cases (22) in Table 6 is only approximately 17% of the total sample, and the number of cases (36) in Table 7 is below a third—hence the result of this second part of the analysis perhaps should be considered a mere replication of those several decades old studies of ‘cumulative advantage’ in science (Allison and Stewart 1974; Reskin 1977; Allison et al. 1982; originally conceptualized by Merton 1968). Another possible analytical angle is of course to declare these extreme cases rather expectable and natural features of the inherently unequal resource distribution in science, which was demonstrated almost a century ago and has been discussed by several prominent scholars (Lotka 1926; Price 1986/ 1963; Cole and Cole 1972), but this clearly lies outside of the scope of this article to discuss in further detail. Within the context of this article’s aim, however, it can rightfully be established as remarkable that individuals that have been awarded a major grant specifically tailored to assist them in establishing their own research line during their postdoc years can emerge as successful in attracting these grants as those listed in Table 6, and not least, that almost a third of the recipients of any one of these grants have previously obtained at least one other grant that either has the very same purpose of providing them with the much-needed research funding in a critical stage of their career that is allegedly not available to them elsewhere, or a highly competitive grant from the Swedish Research Council, or both. This is an unintended but quite expectable effect, that can be (partly) explained by the inherently selective nature of competitive funding, the pluralistic character of the Swedish research system and its lack of coordination, and the forces of institutional isomorphism that appears to accentuate these tendencies. 5. Concluding discussion Underpinning the argument in this article is the recurrently conveyed observation that the Swedish science funding system is pluralist and diverse, with a lack of coordination between different bodies and agencies. While ideally such a system would bring diversity also on the performer side and a healthy mix of different funding programs with different criteria and different aims, experiences of the past years seem to indicate a certain homogeneity in the collection of funding programs. In combination with lack of coordination as well as the limited size of the ‘market’ of possible grantees in Sweden (see below), this homogeneity apparently creates agglomeration effects of the type that have been displayed by the data analysis in a previous section. As stated in the introduction and as shown in the above discussions, the argument has two parts that deal with two levels of analysis. First, we have displayed a striking uniformity in aims and purposes of the funding programs under study, and this we have explained by the use of the concept of institutional isomorphism from neoinstitutional theory. Especially, the two mechanisms of normative pressure and mimesis are typically at work when non-mainstream organizational actors seek to establish legitimacy; mimesis most typically also as a response to environmental insecurity, which in this case is represented by internationalization (especially Europeanization) and not least uncertainty with respect to missions and purposes of public research funding. Several domestic Swedish funding bodies are not only experiencing this uncertainty but also already situated in an incoherent and pluralist system, and will thus, according to the theory, look at role model organizations as well as attempt to more generally adhere to dominating discursive themes to find clues on how to increase their legitimacy. Second, we have shown one consequence of homogeneity in a particular case, namely agglomeration effects. They are most certainly unintended, given the stated purposes of all these programs, but they are simultaneously expectable, since research funding is inherently selective and coordination between different funding bodies is typically weak. Interestingly, however, the agglomeration effects are especially accentuated in Sweden due to the forces of institutional isomorphism which are strong due to the diversification and the recent restructuring which has created stronger uncertainty, among other things by abolishing established disciplinary-oriented organizations with new mission-oriented ones. But some points remain to be discussed. While agglomeration effects in the research (funding) system are neither unwanted nor inappropriate per se but expectable and well-documented in previous studies of the research system (see above), in this particular case they run squarely against the purpose(s) of the programs under study, which is to provide young researchers with the opportunity to build their own research lines. By nature, this purpose would presuppose that there is a relative lack of funding opportunities for young researchers who want to establish their own research lines. This is evidently not the case; the programs under study were launched one by one over a period of several years, and thus it is rather unreasonable to suggest that more than just the first one was conceived with the purpose of filling an urgent gap in the system. Of course, one funding program might not be considered sufficient to solve a structural problem across a whole public science system, and so it can be argued that there should be room for several similar funding programs in Sweden. The analysis here, however, suggests otherwise, not least perhaps the extreme cases noted above and the series of similar grants by different funders secured by some of them (Tables 6 and 7). Regardless of whether the reason for the emergence of these extreme cases is a limited market (in which case Swedish research performers should indeed consider themselves lucky; because this would mean that there is an oversaturation of funding) or an effect of rather natural cumulative advantage in science or an expression of Lotka’s Law (see above), they are doubtlessly examples of excessive resource agglomeration. It is clearly not in our interest here to make normative statements about the purposefulness or rightfulness of such resource agglomeration, only to note that it is expectable (for reasons discussed, see also below) and unintended, because it clearly goes against the stated purposes of the funding programs. Also related to cumulative advantage and Lotka’s Law, research funding is indeed fundamentally elitist in the sense that the paramount ambition (in some cases the only ambition) of funding agencies is to fund science of as high quality as possible. (Although the exact definition of quality is quite evasive, it can be rightfully stated that it is usually similar across programs and funding agencies.) It is normally not part of the mission of a funding agency to weigh in the decisions of other funding bodies when assessing an application and consider it a merit that an applicant has not received (similar) grants elsewhere, since this would logically mean that part of their aim would be to fund research or researchers that nobody else wants to fund. Nonetheless, it should perhaps be noted here that one interpretation of the actual stated purpose of the funding programs under study to solve a structural deficit in the system is exactly that: to fund research or researchers that nobody else wants to fund. We have shown that there are clear and demonstrable inadvertent agglomeration effects in the collection of these currently so popular funding programs for promising young researchers in Sweden. While this might lead some to declare the outcome of the programs unfortunate or even counterproductive given the evidence that such effects have harmful consequences for the careers of those experiencing the cumulative disadvantage that unavoidably is part of it (Melin and Danell 2006), our conclusion is another. The agglomeration effects run counter to the ideas behind the funding programs themselves, but they are also expectable and quite natural, in the sense that unequal resource distribution is an inherent feature of the research system, as has been demonstrated numerous times. The agglomeration effects can therefore not be blamed on the funding bodies, as these are deeply embedded in an inherently unequal system and rely on a paramount quality criterion for research funding that is achieved through competition, that also prevents them from actively working to counter unequal resource distribution, regardless of whether this is among the stated purposes of any of their specific funding programs. We have therefore also contributed by elevating the explanation to the level of the field, as it cannot seemingly be reduced to a blame game among the organizations. Neoinstitutional theory helps explaining why it is that these funding bodies (both public and private) establish so similar funding programs in spite of the lack of coordination in the field, their mutual competition, and their asserted ambitions of creating their own market niche in a pluralist system. An additional (and auxiliary) point can be made. The agglomeration as such may indeed make it seem as though this new fashion in research funding to direct programs toward individuals rather than environments in practice is just a reincarnation of the former. When these successful young individuals manage to secure several different grants, as in the extreme cases above, the funding will unavoidably have to be redistributed to a surrounding group of scientific personnel with the individual remaining in charge, as principal investigator, and so the result will be another excellence environment likely well on par with those previously supported by targeted programs. We have shown with a rough calculation (footnote 7 above) that in some cases, the level of funding will actually also be similar. Which brings us to research desiderata. Two main questions linger and were not within the aims of this article to answer. First, it would be interesting to conduct a follow up study including an in-depth investigation of the correlation (or lack thereof) between the purposes and aims of the various programs and their actual outcomes. Second, a potentially very rewarding inquiry lies ahead for those interested in conducting a qualitative, in-depth, study of a few of the individuals represented in this sample, perhaps most of all (some or all of) the 22 extreme cases and how they manage to spend all the money they have attracted, but also generally (and with possible inclusion of a comparative element) what it actually means for the career prospects of a young researcher to be identified by a research funding organization as a ‘future research leader’. Notes 1. The ‘wage earner’s funds’ was a project launched by the social-democratic government in the early 1980s, after years of agonizing internal and external debate, that constituted the first step toward a socialization of Swedish enterprise by the forced gradual transfer of stocks from companies to governmentally owned ‘wage earner’s funds’. After a change of government in the early 1990s, a new center-right government abolished the funds and the accumulated capital was invested in the public pension funds as well as used to create 10 public research foundations; these 10 shared a total of over SEK 10 billion (E1.13 billion) (Benner and So¨ rlin 2007: 34). 2. The Humanities and Social Sciences Research Council, the Medical Sciences Research Council, the Natural Sciences Research Council and the Technical Sciences Research Council. 3. The Swedish Council for Working Life and Social Research and the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning. 4. The figures on funding in this paragraph and in Table 1 are all from the public so-called Statistical Database on Swedish Higher Education Institutions (Statistikdatabas om Ho¨gskolan), maintained and provided online by the Swedish Higher Education Authority (Universitetskanslera¨mbetet), which collects the data annually directly from the universities and colleges and from secondary sources such as Statistics Sweden (http://www.uk-ambetet.se/ statistikuppfoljning/statistikdatabasomhogskolan). 5. In the years covered here (2008–11), acceptance rates for grants within the Swedish Research Council’s general annual call have oscillated between 19 and 23% (over the whole disciplinary spectrum). 6. To achieve such a complete mapping of individuals’ funding portfolios, a search of project and grant databases with individuals’ names as search strings would not suffice, as several likely funding sources and funding streams are more or less hidden, i.e. not announced in detail online or in funders’ annual reports, not to mention non-monetary resources disposed by university researchers as well as questions of whom to include in a person’s research group, and so on. 7. Coincidentally and interestingly, this is a sum of money well on par with those grants previously awarded to groups and environments within the government’s flagship excellence funding program, the Linnaeus Grants (Hallonsten and Silander 2012: 374). 8. The table has been stripped of redundant information: Since they only make up background information, VR and ERC grants in 2012 have been omitted altogether, and omitted in all cases where they show up the same year as, or after, SSF or ICA grants and no KAW grant has been awarded in 2012.

It is widely acknowledged that research collaboration is a key mechanism for both knowledge production and diffusion in science and technology (Steensma, 1996; Ahuja, 2000; Hagedoorn et al., 2000; Powell et al., 2005). However, we know little about the institutional factors that influence the capability of public research systems to connect distributed knowledge and competencies across institutional and organizational boundaries. While the institutional interfaces between university and private sector research are comparatively well understood (Meyer-Krahmer and Schmoch, 1998; Owen-Smith et al., 2002; Calvert and Patel, 2003), there is a lack of systematic and comparative knowledge as to the institutional conditions that facilitate external work relations between public research organizations. A burgeoning literature addresses either the individual and network level of collaborative research (Landry and Amara, 1998; Melin, 2000; Newman, 2004; Lee and Bozeman, 2005), or the growth of international scientific collaboration (Georghiou, 1998; Gl¨anzel, 2001; Jappe, 2007). Among the few studies that deal with the role of institutional structures in the formation of research collaboration in public research are Laudel and Gl¨aser (1998) who examine the boundary-spanning role of collaborative research centers, and Corley et al.’s study of epistemic and organizational institutionalization in two large-scale, multi-discipline collaboration programs (Corley et al., 2006). Yet the question of how governance structures of public research organizations influence extramural collaborative research has rarely been addressed in recent sociology of science and public policy studies. The desire to know more about the factors that contribute to research collaboration is given further impetus by the substantial changes seen over the last three decades in the institutional and organizational conditions under which scientific research is conducted (Senker, 2006; Shapira and Kuhlmann, 2003; Gornitzka et al., 1998). The need for effective inter-institutional knowledge flows is particularly critical in emerging research domains on the one hand, and institutionally differentiated research systems on the other hand. In emerging science and technology domains, a sizeable share of research is conducted at the intersection of established scientific disciplines and across fundamental and applied research, seeking for and building on cognitive and institutional complementarity. One such research domain is nanoscale science and technology, referred to as “nano S&T” in the following. Nano S&T embraces several disciplines and research areas, such as applied physics, materials science, physical chemistry, physics of condensed matter, biochemistry and engineering and polymer science, and potential application areas as diverse as drug delivery, environmental sensing, manufacturing, and quantum computing (Porter et al., 2008; Heinze, 2004; Heinze, 2006; Hullmann and Meyer, 2003). The need for effective inter-institutional knowledge flows is also critical in public research systems with a high level of institutional differentiation. From an innovation policy point of view, such research systems need not only to allow knowledge diffusion across institutional boundaries via career paths, but also to institutionalize effective mechanisms to support day-today collaborations across organizations that scientists seek to establish and maintain. In this regard, the German research system (GRS) is an interesting example due to its high level of institutional fragmentation. In addition to the universities, the GRS embraces a large extra-university sector including institutes of the Helmholtz Research Centers (HGF), the Max-Planck Society (MPG), the Leibniz Association (WGL), and the Fraunhofer Society (FhG). These organisations have developed quasi-functional monopolies in particular research domains, such as fundamental research (MPG), applied contract research (FhG), and big-science research facility management (HGF). In consequence, they have traditionally not collaborated much with each other (Hohn and Schimank, 1990). Therefore, studying the institutionally fragmented GRS in the emerging domain of nano S&T is an interesting case, because it touches upon the question how well the differentiated institutional structure of the GRS is aligned to the need for inter-institutional research in the emergent research domain of nano S&T. If institutional segmentation is viewed as an obstacle for effective knowledge exchange, then systems with such structures in their public research system are not expected to be among the top performers internationally. This, however, is not true for German nano S&T research. The GRS scores relatively high in what has been called the “global nanotechnology race”. In absolute terms, Germany ranks fourth in publication output worldwide and third in patent applications at the European Patent Office (Hullmann, 2006). In relation to GDP measures, Germany ranks even higher (Heinze, 2004: Fig. 4). Our analysis addresses three interrelated questions. First, how common is research collaboration across institutional boundaries empirically? Second, what are scientists’ rationales to engage in collaborative research? Third, which governance structures are either conducive to inter-institutional research collaboration or interfere with scientists’ efforts to engage in collaborative work? For answering these questions, we frame our analysis by dimensions of governance that reflect the ongoing debate on the coordination of autonomous, but interdependent actors (Hollingsworth and Boyer, 1997; L¨utz, 2003). Furthermore, we draw on multiple data sources such as annual reports of German research institutions, internal reports and communications, co-publication analyses, and macro research statistics. Most importantly, we conducted 32 semi-structured interviews between 2004 and 2006 with representatives of all non-university research organizations, the German Ministry for Education and Research (BMBF), institute directors at universities and extra-university institutions, and senior researchers and junior group leaders in the field of nano S&T. More details on our data are presented in Appendix A. Sections 2 and 3 provide key facts about the domain of nano S&T and the GRS. Sections 4 and 5 sketch current collaborative activities in nano S&T across research institutions in Germany and discuss rationales for cooperative research relationships across institutional boundaries. In Sections 6 and 7, we elaborate on institutional factors that either facilitate or hinder the transfer of knowledge and expertise between research organizations. Section 8 summarizes our findings and gives an outlook on research desiderata. 2. Nano S&T: research across established cognitive boundaries Nano S&T is one of the most thriving research domains worldwide. Recent data show that the num ber of worldwide scientific publications in nano S&T has increased by a factor of six in the past ten years (Hullmann, 2006). Likewise, the annual number of patent applications in nano S&T at the European Patent Office more than doubled in the last decade (Scheu et al., 2006). Worldwide public funding has increased from D 500 million in 1999 to D 385.000 million in 2004 of which the United States, Japan, the European Commission and Germany together invested about 70% (European Commission, 2005). One of the conspicuous characteristics of nano S&T research is its high level of research across established disciplinary and field boundaries, but also across the traditional distinction of basic and applied science. Although the cross-disciplinary character of nano S&T is sometimes questioned (e.g., Schummer, 2004), several studies have shown that this research domain shows a remarkable degree of research activities that cut across established cognitive boundaries. This means that a sizeable share of research is conducted at the intersection of established scientific disciplines and across fundamental and applied research. Among the earlier studies, Braun and Meyer (1998), based on bibliometric measures, identified nano S&T research as more cross-disciplinary than science in general. More recently, Heinze and Bauer (2007), based on a longitudinal multimethod research design, show that one key explanatory factor for research creativity in nano S&T is the ability of scientists to effectively communicate with their colleagues and their capability to address a broader than average work spectrum. Furthermore, Rafols and Meyer (2007), relying on both interview and bibliometric data, find a consistently high degree of cross-disciplinarity in the cognitive practises of scientists in the field of molecular motors. The authors argue that the need for a broad set of research instrumentalities, such as fluorescent microscopy or X-ray crystallography, is one of the main drivers of boundary crossing research in nano S&T. Their findings are in line with science history studies on the pivotal role of research instrumentalities as connectors between independent research specialties and disciplines (Shinn and Joerges, 2002). Among the various nano S&T subfields, our analysis focuses on nano-electronics and nano-interfaces. Nanoelectronics is an emerging subfield with topical areas, such as carbon nanotubes or wafer bonding. Carbon nanotubes have interesting electrical properties that are scientifically relevant for molecular electronics and biophysics; at the same time, however, carbon nanotubes have a high potential for future integrated circuits and thus for the computer industry.Wafer bonding is another nano-electronical area where epitaxy (method of thinfilm deposition) methods are used to allowfaster electron transmission within silicon structures, a development that is highly relevant to enhancing computer processor speed. Nanoscale interfaces is a second emerging field within nano S&T, spanning topical sub-areas such as nano-capsules or nano-sensors. Based on thin film colloidal chemistry methods, nano-capsules have considerable potential to be used as carriers for targeted medication. Similarly, the fundamental understanding of the reactivity of nano-surfaces allows the construction of biocompatible and portable nano-sensors. 3. The German research system: key facts and recent dynamics Before examining collaboration in nano S&T within the GRS, we introduce some key facts on its institutional structure and recent dynamics. A striking feature of the GRS is the relatively large share of extra-university public sector research. Comparing input and output variables shows that the German university sector is larger than the extra-university sector in terms of personnel (Table 1, B), but has a much smaller research budget per researcher (Table 1, C). Nevertheless, university researchers are highly productive, as displayed by their share in all three output categories (Table 1, E–G). Within the extra-university sector, the MPG has the strongest scientific profile. While MPG institutes recruit only 4% of German research personnel in the natural sciences (column B), they account for 10% of the German Science Citation Index (SCI) papers (column E) and 34% of all German Science and Nature articles (column F). In contrast, FhG institutes publish much less in the SCI but have the highest relative output of patent applications (column G). FhG institutes primarily conduct contract research for companies, but also for public agencies. Their core funding is substantially lower than that of all other research institutions (column D). In terms of research output, universities are located in between the distinct institutional profiles of MPG and FhG. The HGF has traditionally had an institutional mission in big science research facilities and nuclear technology development and thus has stronger ties to the federal state. Although similar to the MPG in its high level of institutional funding (Table 1, D) and equipment level per researcher (Table 1, C), its relative productivity is substantially lower: 11% of German research personnel in the natural sciences (Table 1, B) publish 7% of the German SCI papers (Table 1, E) and 14% of all German Science and Nature articles (Table 1, F), and file 13% of all patent applications among the public sec tor research institutions (Table 1, G). WGL institutes are also an important part of the German research landscape. Their overall relative research performance (4% of SCI publications) matches their relative size (3% of research personnel). However, theWGLhas not developed a clear institutional profile on the upper organizational level thus far. Fig. 1 maps the research profiles of German research institutions on two major output variables: publications and patent applications, both relative to 100 R&D staff between 1991 and 2002. These highly aggregated dimensions are useful to locate various profiles in the German public research system. Basic science (upper left area) and technology-driven research (lower right area) are positions occupied by the MPG and the FhG, respectively. Two trends are discernible in Fig. 1. First, all institutions substantially increased their productivity between 1990 and 2002, as is visible by the movement both towards the right and upwards, indicating higher outputs per R&D staff. These shifts are a clear indication of the increasing pressure on the research system to demonstrate higher output efficiency. In the same period, public sector research funding decreased substantially: between 1990 and 2002, funding, particularly of the university sector, decreased by about 10% in real terms. In addition, the number of tenured university professors decreased from 25,000 to 23,000 between 1995 and 2005, while the scientific labor force in the public research sector stagnated (DHV Press Release 11/2005; BMBF, 2005: Tables 20, 21, 38). Despite this decrease in funding, scientists produced significantly more research papers and patent applications in 2002 than in 1990. Second, while Fig. 1 does not indicate fundamental changes in the relative positions of German research institutions one should note, however, that current pressures on the research system have induced competition between formerly protected research domains. Shifts of research organizations in the direction of technological research (shift to the right in Fig. 1) tend to be more pronounced than movements in the direction of scientific output (upward movement in Fig. 1). Institutes that did not conduct technological research in the early 1990s apparently do so today. It also implies that institutes whose core competence has traditionally been in technology research have come under considerable pressure. Consequently, the FhG – financed largely through contract research with industry – today faces increasing resource competition from other research institutions. In sum, the problem of institutional segmentation and dominance of organizational self-interests, as observed by a high-level evaluation committee (Brook et al., 1999), is still highly relevant. Budget cuts and pressure on output efficiency have increased competition for research funds, and might hamper synergies within the GRS. 4. Collaboration across institutional boundaries in Nano S&T In this section,weaddress the first question: howcommon is inter-institutional research collaboration in the GRS in the field of nano S&T? In order to measure the level of inter-institutional collaboration, we systematically searched for collaborative research activities in the two subfields of nano-electronics and nano-interfaces. First of all, we identified nano-related publications and collaborative research projects by comprehensive search strategies. Further, we conducted interviews primarily with researchers who were experienced in extramural collaborations, but also with scientists who reported few external contacts only. This led to the identification of further types of collaborative activities. Using co-publications as a bibliometric indicator (Melin and Persson, 1996; Bordons and G´omez, 2000; Gl¨anzel and Schubert, 2004; Newman, 2004), we find that - at the level of research organizations - the majority of research collaborations are observed between universities and the extra-university research sector, while co-authorship relationships between organizations of the extra-university sector are tenuous. The MPG collaborates most frequently with universities, followed by the HGF and the FhG, but there are few co-publications between MPG, FhG or HGF (Table 2). These results confirm the conclusions of the Brooks report which criticized the low level of collaboration within the extrauniversity public research sector (Brook et al., 1999). At the level of institutes within universities and extra-university research organizations, we identified formal project collaborations by a systematic screening of research projects funded either by the German Research Foundation (DFG) or the Federal Ministry for Education and Research (BMBF). The DFG has been funding basic research projects in the areas of nano-colloids and -polymers, nano-materials and optical nano-technologies. These programmes have been extended in size and scope over the last decade and thus provided ample opportunities for collaborative activities to develop. In the applied research funding of BMBF, we found collaborations in the fields of nano-polymers, semiconductors, nano-materials and laser. Some of these projects are part of the two broad subfields mentioned in the above, and several scientists from such projects were selected for interview. Interviews helped identify other types of formal collaborations. There are, for instance, cooperation contracts between research institutes specifying use of research instrumentation and interchange of personnel. We identified junior research groups at the intersection of institutes that were located at one institution, but personnel and instrumentation costs were shared among two or more institutions. Furthermore, education of junior researchers is an institutional vehicle for collaborations, not only between universities (where junior staff receive their doctoral degrees) and the extra-university sector (where they carry out their projects), but also within the extra-university sector. In addition, various forms of informal collaborations exist, including meetings of the heads of institutes whose function is information sharing and preparation of collaborative research proposals; also, doctoral students who travel between sites and carry out experiments are shared. In sum, the interview data suggests a more nuanced picture of collaborative relations, particularly within the extrauniversity research sector. Our interview data confirm the conclusion of earlier literature that co-publications are only a partial measure of collaborative activities (Katz and Martin, 1997; Laudel, 2002). 5. Rationales for research collaboration Understanding research collaboration in a highly differentiated research system requires consideration of scientists’ rationales for engaging in collaborative activities. Generic motives for research collaboration include curiosity, knowledge advancement, sharing the excitement of a research area with other scientists, or intellectual companionship (Katz and Martin, 1997; Beaver, 2001). These motives are anchored in what Luhmann (1975) describes as a “cognitive style of expectation”. However, these motives do not specify why particular scientists would collaborate at a given time. For the field of nano S&T, we empirically validated the following additional collaboration rationales. The first set of rationales is expansion of research capacity, which embraces (a) the need for complementary knowledge and expertise; (b) access to equipment and instrumentation; and (c) the ability to build consortia that compete for funding. An example for (a) is an ongoing collaboration between two groups, one of which specializes in the electrical measurement of nanowire characteristics, while the other is highly knowledgeable in respective optical measuring techniques. Both knowledge domains have been fruitfully combined over time, leading to many co-authored publications. Combining complementary knowledge and expertise expanded both groups’ capacities to address new questions and to enter new thematic areas. An example for (b) is one group interested in solving a particular research question on metallic nanoparticles and two instrumentation groups (synchrotron and molecular beam lithography) that are interested in learning more about the various possibilities of their complex machinery. There were many examples for (c). Because expanding research capacity requires additional funding and many research questions (due to their complexity) cannot be addressed by single groups alone, researchers have an incentive to build project consortia that compete collectively for thirdparty funds. A second set of collaboration rationales is anchored in strategies to improve current research capabilities. It includes (d) keeping own research activities focused and (e) learning new skills or techniques. Examples for (d) and (e) are three chemistry groups that are embedded in institutes with strong physics capacities. Such embedding has several benefits: the most important are access to new research questions generated outside a given specialty and opportunities to become acquainted with new methods and instrumentation, but also important is continuous scrutiny from the physicists with regard to interpretation of experimental results. Third, realisation of institutional complementarities is an important collaboration rationale: Universities seek cooperative relations with extra-university institutes to obtain access to facilities, instrumentation, and research topics, while extra-university institutes depend on access to students and junior researchers. Institutional complementarities also exist between groups specialized either in basic or applied research. FhG institutes usually provide considerable expertise in testing and development of reliable technical processes, while university or MPG groups have access to the latest knowledge at research frontiers. In the areas of nano-electronics and nanointerfaces, such institutional profiles have been found to be complementary for both sides. On the one hand, there are novel scientific approaches in wafer bonding and nano-polymers that require considerable engineering before their industrial application becomes feasible. On the other hand, problem solving on the engineering side has generated new research questions that are valuable for a fundamental science perspective. Fourth, research institutions seek collaborations to enhance their visibility for scientists and companies in the field. We identified cases in which collaborators related to each other because of their different research profiles that in turn are anchored in different organizational missions. There are MPG institutes (not the majority, however) that use their FhG collaborations to signal to industrial companies their openness to applied technological research questions (which traditionally lie outside their core competency). Contacts with larger companies can be beneficial for MPG institutes in terms of additional funding, but they also have value with regard to future job opportunities for doctoral students and post-docs. Conversely, a number of FhG institutes (not the majority, however) employ contacts with MPG institutes to signal scientific prestige to academic researchers in university departments and other basic science facilities. Furthermore, because the FhG funding regime allows only little exploratory research, such contacts signal access to research frontiers, which, in combination with engineering and reliability testing capacities, might be an incentive for companies to fund contract research in FhG institutes. The difference between the MPG and FhG institutes is that the former use signalling primarily to attract industrial recognition, while the latter attempt to draw either academic or industrial attention to their research activities. 6. Factors conducive to inter-institutional research collaboration Rationales for research collaboration across institutional boundaries lead us to the third question: which governance structures are conducive to inter-institutional research collaboration? Recent publications on the GRS investigate primarily interdisciplinary cooperation (R¨obbecke et al., 2004; Lengwiler, 2006), and only a few studies deal in more detail with the institutional framework of the GRS, but they discuss data from the 1980s and early 1990s and do not cover more recent institutional developments (Hohn and Schimank, 1990; Hohn, 1998; Laudel and Gl¨aser, 1998). In order to examine the governance of interinstitutional research collaboration in more detail, we refer to a governance cube as a heuristic tool (Fig. 2). Generally speaking, governance refers to analytically distinguishable forms of institutional coordination of autonomous, but interdependent actors. Hierarchy, competition, network, association, and community are ideal types of governance capturing the rules of a game at a highly generalized level (Hollingsworth and Boyer, 1997; L¨utz, 2003). In reality, these governance forms are often interconnected, thus forming governance regimes. Benz (2007) argues, for instance, that actors have to find out how to cooperate with competitors or to compete with partners in networks, to negotiate an agreement under tight organizational constraints, or to find approval for the outcome in external arenas in their own organization or group. The governance cube takes up notions of both governance forms and governance regimes but is specifically tailored to the research question of interinstitutional research collaboration. The dimension of thematic interdependence captures the extent to which research activities build on each other and how the cognitive structure of research fields impinges upon the work organization of research. The organizational dimension depicts the governance regimes of both the university and the extra-university sector including HGF, MPG, and FhG. On the level of single research units (institutes, groups), the organizational dimension embraces variables such as internal differentiation, permeability of communication across levels of hierarchy, career incentives, or research missions. Resource endowment includes the quantity and the quality of staff and equipment as well as the funding structure of research units (Fig. 2). By applying dimensions of the governance cube, we identified a number of institutional factors that are important in facilitating inter-institutional research cooperation in the GRS. As far as the intellectual dimension is concerned, specific thematic profiles of research groups (and research institutes) are of paramount importance because they support search processes and decision-making (ex ante), and increase mutual benefits from collaborative activities (ex post). Many of the interviewed groups tend to be highly interdependent both in terms of the interdisciplinary character of work and also with regard to the need for complex instrumentation and materials. This point is in accordance with the finding that one of their major rationales for collaborative activities is the need for complementary knowledge and expertise. It also fits our finding that researchers prefer collaborators with a reputation for a certain expertise that proves valuable in research consortia competition for additional research grants (see Section 4). Specific profiles are also important with respect to the organizational dimension, but here they pertain to the “research mission” of groups or institutes. Such specific research profiles include basic versus technology-driven research, the capability to conduct highly reliable routine research or the capacity to conduct research at scientific frontiers. Organizational and intellectual profiles need not overlap. Further along the organizational dimension, recruiting qualified research personnel with a record of job mobility endow the employing organizations with a better understanding of different institutional perspectives. This organizational capacity seems valuable in a functionally differentiated research system, as is the case in Germany. Researchers with inter-organizational career tracks or with a record of visiting fellowships enable informal contacts to other research institutions that help in building consortia at certain times and for particular purposes. In addition, research leadership facilitates collaborative activities across institutional borders. Research leadership means conceiving and implementing midterm research goals which enable external coalition building. It also means proactive strategies to access external funding and the ability to shift the initial research goals in the direction the research is moving. Research leadership is in accordance with the rationales of expansion of research capacity and improvement of current research. Finally, effective administration at the organizational level supports research collaboration, for instance, by making decisions promptly, by not consuming resources above a certain threshold (“overhead”), or by allowing flexible interchange of resources, including mobility of personnel. With regard to the resource endowment, our analyses suggest that research collaboration is facilitated when partners have sufficient core funding at the group or organizational level. Such funding is obviously a prerequisite for developing specific research profiles, which support search processes and increase mutual benefits from research collaboration. Findings from our interviews also suggest that sufficient core funding is a prerequisite for engaging in research venues that are intrinsically risky, a finding that pertains in particular to research creativity. However, third-party funding also stimulates cooperative behaviour; external collaboration is requested in many funding programmes. One of the major benefits of third-party funding is that it helps research groups keep their research focused and coordinate various research agendas. Institutes with a high level of core funds compete for third-party funding only if the research leadership decides to do so. MPG and HGF departments, for instance, which traditionally enjoy very high levels of core funding (Table 1), tend to be less involved in extramural collaborative research projects if their research leaders do not actively seek third-party funding. Neither core funding nor third-party funding alone induce collaborative activities in the field of nano S&T. Instead, a balance between the two seems advantageous. Furthermore, resource flexibility appears to be important in facilitating extramural research collaborations. Flexible allocation and interchange of resources between institutes supports collaborative activities because this flexibility helps to conduct research effectively. One example is scientists who, while changing jobs from one institution to another, take their research projects with them. Another example is the shifting of project funding from aMPGinstitute to a university institute because a collaborating doctoral student has access to special equipment at the university and thus can carry out the work more effectively. A third example is collaboration contracts between extra-university institutes arranging mutual support in instrumentation or library services. 7. Barriers to inter-institutional research collaboration With regard to the organizational dimension, stereotypes and prejudices tend to impede cooperation between various research organizations. Examples of stereotypes that we validated in our interviews are as follows: HGF researchers have a reputation for being slower and less productive than average, while MPG scientists are viewed as those with lavish laboratories and sometimes arrogant attitudes towards researchers from other research organisations. In contrast, FhG researchers are often equated with industry because they primarily focus on money instead of scientific quality. Furthermore, university researchers are often regarded as conducting research projects in a chaotic and even unprofessional way. These examples are not based on experience, but rather on hearsay, because both low overall job mobility and a low degree of formal and informal inter-institutional collaborations have provided only limited opportunities for genuine experiences with other research organizations. Second, and in contrast to the first factor, interinstitutional collaboration can be hampered by incompatible working routines anchored in divergent organizational missions. Interviewees from FhG institutes and MPG institutes agreed in their assessment that straightforward interaction between what they called the “engineering attitude” of FhG researchers (i.e. to produce a project result within a finite time frame and with a finite sum of money) and the “playing attitude” ofMPG researchers (i.e. searching without restrictions or “picking flowers”) can be bothersome if there is no facilitator or translator. Combining divergent working routines in a synergetic fashion requires mobility record and/or active research leadership at the level of institute directors. Third, lack of interface management seems a common problem for researchers who do not dispose of means or resources to organise follow-up activities in cases when they have results that might be relevant for other research institutions. It was only very recently that the headquarters of the MPG and the FhG started a dialogue on pooling expertise and know-how in various research areas, among them nano S&T (Gruss, 2002: pp. 19–20). Regarding the resource endowment dimension in the governance cube (Fig. 2), our analysis suggests that sustained budget cuts over the last decade, particularly in the university system, have negatively affected the ability of research groups to engage in inter-institutional collaboration. This situation has been counterbalanced only partly by the comparatively good funding situation in the field of nano S&T. Immediate effects of funding restrictions are the discontinuation of ongoing cooperation or the loss of future options for collaboration. These impacts pertain especially to the university system, where research collaboration covered by core funding has become difficult over time. We have argued above that combinations of core and third-party funding together provide incentives to build up specific research profiles and seek extramural collaboration. Such a mix seems advantageous, compared to either mere core or project funding. However, if core funding falls below a certain threshold, capacities for building and sustaining research profiles will decline significantly which, in turn, inhibits the search for collaboration partners and the opportunity to gain from collaborative activities. These results are corroborated by Laudel (2006). Mid-term effects of funding cuts include the emergence of status hierarchies between the university and the extra-university sector. Table 1 shows thatMPGinstitutes have a budget/research personnel ratio of 0.264, while this ratio for the universities is merely 0.086. Thus, according to this coefficient, MPG researchers are about three times better equipped than their university colleagues. This finding is consistent with our interview results showing that university researchers increasingly experience problems in catching up with the instrumentation and research equipment of MPG institutes; thus, they are not well-positioned as research partners. However, apart from budget restrictions, accompanying regulatory structures also have adverse effects. First, research careers have become increasingly unattractive: not only have real income opportunities for younger researchers been levelled downward, but also current changes in labor law have, in fact, erected new barriers to job mobility because researchers face real income (or pension scheme) losses when moving from one type of institution to another. Second, budget cuts have been accompanied by New Public Management (NPM) reforms that substitute hierarchical for academic control (Boer et al., 2007). In his analysis of such NPM reforms in the United Kingdom, Georghiou (2001: p. 294) argues that public research sector institutions have been converging in their research activities and profiles, thus narrowing the capabilities of the research system as a whole. 8. Discussion and conclusion Our analysis started with the observation that the need for effective inter-institutional knowledge flows is critical in the emerging domain of nano S&T, where a sizeable share of research is conducted at the intersection of established scientific disciplines and across domains of fundamental and applied research. In addition, organizing effective knowledge flows is a particular challenge in institutionally segmented research systems, such as the GRS with a broad and differentiated extra-university research landscape. However, since little is known about the factors that influence the capability of public research systems to connect distributed knowledge and competencies across institutional and organizational boundaries, we investigated inter-institutional knowledge flows within the GRS in the domain of nano S&T. First, we find that the majority of domestic research collaborations are observed between universities and the extra-university research sector, while co-authorship ties within the extra-university sector are tenuous. Our qualitative data suggests a more nuanced picture of collaborative relations including cooperation contracts between research institutes, joint junior research groups, or informal meetings of institute directors, particularly within the extra-university research sector. Second, we find that scientists collaborate primarily to expand and improve their research capacity, to benefit from institutional complementarities, and to enhance their visibility within the research field. Our findings add to collaboration motives identified by previous literature including curiosity, knowledge advancement, sharing the excitement of a research area with other scientists, or intellectual companionship (Katz and Martin, 1997; Beaver, 2001). Third, we identify specific thematic profiles, recruitment of research staff, support for job mobility, research leadership, balanced core and third-party funding, and flexible mechanisms for funding allocation as institutional conditions that are conducive to inter-institutional research collaboration. In contrast, organizational stereotypes and prejudices, incompatible working routines anchored in diverse organizational missions, lack of interface management, and sustained budget cuts particularly in the university system, have had negative impacts on scientists’ opportunities to engage in collaborative work relations outside their home institution. Our empirical evidence suggests that the institutional structure of the GRS is permeable enough to allow sufficient knowledge flows in the emergent research domain of nano S&T between universities and the extrauniversity research sector. In fact, this seems one key explanation why Germany is among the top players in the global nanotechnology race. In recent years, the Max-Planck-Society and the Helmholtz-Association have taken several measures to improve collaborative relationships with universities, for instance via the establishment of newjoint junior research groups, or contracts with universities that allow mutual access to instrumentation and library services, and the establishment of new research schools. Sufficient knowledge flows cannot be observed, however, between organizations of the extra-university public research sector. Here many of the adverse effects of the segmented institutional structure, described in Section 6, are salient. The high level of segmentation becomes especially obstructive when, like in the case of the GRS, research systems operate both under output pressure and resource stagnation. Recent efforts to establish stronger collaborative relationships between the MPG and the FhG are exceptional and are not indicative of a paradigm shift within the public non-university research sector. The strong position of German research in global nano S&T has also to do with the many collaborative ties to research labs abroad. Several of the studied research teams are active collaborators of groups in Europe, the United States and Russia. One key condition of these collaborations could be the gradual emergence of the “European Research Area” (Kuhlmann, 2001), but still only little is known about the institutional conditions (organizational cultures, funding systems, intellectual property rights regulations, career paths, or promotion criteria) for effective knowledge transfer in the public research sector (in nano S&T) across inherited national research systems—a field for future research. Acknowledgement This paper is based on research sponsored by the Deutsche Forschungsgemeinschaft (FOR 517). The authors gratefully acknowledge helpful comments on earlier versions of this paper from Dorothea Jansen, Uwe Schimank, Harry de Boer, J¨urgen Enders, and one anonymous reviewer. Research assistance was provided by Martin U¨ belho¨r, Christian Pangert und Rebecca Rangnow. Appendix A. Interview data We conducted, in total, 32 semi-structured interviews between 2004 and 2006 with representatives of all non-university research organizations (except for WGL), the German Ministry for Education and Research (BMBF), institute directors at universities and extrauniversity institutions, and senior researchers and junior group leaders in the field of nano S&T. We conducted interviews primarily with researchers who were experienced in extramural collaborations, as displayed in their number of external project collaborations. Interviewees with few external contacts were also included. The average length of interviews is ca. 1.5 h. Interviews were fully transcribed and coded into dimensions and factors.

Measuring gender differences in research and teaching productivity has been a topic of interest for researchers and decision makers for many decades [21]. The basic generalization found in the literature is that male faculty outperform female faculty [5, 10] (for a more comprehensive list, please see [21] and references therein). However, more recent studies have shown that this is by no means due to some inherent superiority of one gender over the other. Xie and Shauman [45] conducted four large, nationally representative, cross-sectional surveys spanning several decades, observing that differences in research productivity declined over the period 1969–1993. Similarly, Gander [21] concludes that, although analysis of data at first glance supports the generalization, further scrutiny reveals that patterns of employment and distribution of funding are actually the main causes of the observed differences in productivity. Once the analysis is adjusted, it is revealed that female faculty have significant research productivity [21]. On the other hand, gender differences in research collaboration have received considerably less research attention in spite of the fact that contemporary scientific research is increasingly conducted in collaboration [2]. Research productivity and collaboration can be studied at various levels: intra-institutional (within an institution), institutional (between institutions), national (within a country), international (between countries), disciplinary (within a scientific discipline), inter-disciplinary (between scientific disciplines), etc. Current Research Information Systems (CRISs) offer rarely exploited opportunities for scientometrics studies at the intra-institutional and national level. Exceptional examples are studies by Leeuwen et al. [42] who performed bibliometric analysis of research performance of a Dutch university using data stored in its institutional CRIS and Perc [36] and Kastrin et. al [26] who investigated research productivity and collaboration of Slovenian researchers relying on publication data recorded in the national CRIS system of Slovenia. In our previous work we proposed a methodology based on co-authorship networks extracted from CRIS databases to analyze intra-institutional research collaboration [38]. In spite of the fact that CRIS systems can be exploited to investigate gender differences in research productivity and collaboration at the intra-institutional level we are not aware of any such study. In this paper we propose a methodology for gender-based analysis of intra-institutional research productivity and collaboration. The methodology is based on statistically robust analysis of enriched co-authorship networks whose nodes have gender labels. Co-authorship networks reflect the structure of research collaboration. Enriched co-authorship networks are co-authorship networks whose nodes are extended by a rich vector of metrics which reflect productivity, collaboration and institutional importance of researchers. The methodology is implemented within a tool called GERBER (GEndeR Based Evaluation of Researchers). We also performed an XML-based integration of the tool with CRIS UNS. CRIS UNS is an information system for storing and managing data about scientific research activity at the University of Novi Sad (UNS), Serbia. It was developed by the recommendations of the non-profit organization euroCRIS [16]. CRIS UNS provides a comprehensive list of publications of researchers affiliated with UNS and enables automated evaluation of UNS researchers and institutions. Using GERBER we performed the gender-based analysis of productivity and collaboration of researchers employed at UNS-PMF – the Faculty of Sciences, University of Novi Sad, Serbia. The rest of the paper is structured as follows. The overview of related research works is given in Section 2. The next section describes our methodology for gender-based analysis of intra-institutional research productivity and collaboration. The accompanying tool is presented in Section 4. The analysis conducted using the GERBER tool is presented in Section 5. Integration of analytic tools with CRIS systems is discussed in Section 6. In the last section we give conclusions and directions for future work. 2. Related work In this section we will review related studies of gender differences in scientific productivity and collaboration (Section 2.1) and give the background of the CRIS UNS information system for storing and managing data about scientific research (Section 2.2). 2.1. Scientometric analysis of gender in research The comprehensive multi-decade study by Xie and Shauman [45], besides observing the decline of gender differences in research productivity, also correlates gender productivity differences with gender differences in personal characteristics, structural positions, and marital status, implying that gender differences in research productivity stem from gender differences in structural locations, and as such respond to the secular improvement of women’s position in science. Contemporary data in the field of social science shows that not only did gender differences disappear in the younger generations of researchers, but that if some differences exist, it is the young female researchers that outperform their male peers [3]. In educational psychology, on the other hand, although females are gaining ground in terms of primary and secondary article authorship and journal editorial board membership, this increase does not keep pace with the male-female ratio in organizational memberships [19]. Also, in the industrial and organizational psychology there are significant gender differences with respect to publication output (fewer publications authored by female researchers) and career courses (male researchers have longer careers, but with longer interruptions) [27]. However, the projected future promises a much more balanced situation in this field [27]. In the domains of science, engineering and technology, the trends are also varying, with the overall impression that participation and performance of women improved in recent times. Within Spanish natural resources and chemistry scientists, no significant differences in productivity were found between genders within professional categories, but the outliers with the highest production were for the most part male [6]. In nano science and technology, female researchers are scarce in number, but perform equally in terms of scientific production and impact [40]. A comprehensive study involving the DBLP database of computer science publications ranging from 1936 to 2010 also indicates a low percentage of women in this field, albeit a steadily rising one [9]. The authors observed that men publish more than women, but attributed this to the fact that the average research life of men is longer. In software engineering journals, the percentage of women authors is roughly on par with the general trend in computer science (16–17%), comparable to the percentage of female editors (18%), but women editors-in-chief are under-represented (9.5%) [43]. Studies of gender differences in scientific productivity also produce varying results in different countries. In Croatia [37], within the studied young research population, females are somewhat less productive than males, with structural variables being the most powerful factor influencing this distinction, which is in line with the observations by Xie and Shauman [45] discussed earlier. In Italy [1], there is also evidence of higher overall male productivity, but with difference smaller than reported in a large part of the literature, confirming an ongoing tendency towards decline. Russia [35], on the other hand, still exhibits strong gender disparity, which can also be said for Turkish social sciences [34]. In Spain, there is no overall significant difference among young postdoctoral researchers in terms of productivity [7], with women leading the way in the number of citations, but also in the number of researchers with no output after obtaining their Ph.D. The above constitutes only a small representative sample of studies dealing with gender differences in scientific production and teaching. A comprehensive meta-study of scientific literature on women in science and higher education, considering almost 1500 articles, shows continued growing interest in the topic, featuring more than 3000 authors, 67 countries, and 86 research areas [12]. Gender differences in research collaboration have received considerably less research attention compared to gender differences in research productivity (an overview of existing studies can be found in [2]). Additionally, the literature shows mixed conclusions about gender aspects of research collaboration. For example, the study by Bozeman and Gaughan [8] from 2011 indicated that U.S. female researchers tend to have more collaborators than U.S. male researchers, while the study by Zeng et al. [46] from 2016 indicated exactly the opposite. The most comprehensive analysis of gender differences in research collaboration at the national level was performed by Abramo et al. [2]. Their study involving the entire population of Italian researchers in the ”hard” sciences and economics showed that gender differences in research collaboration do not exist at the intra-institutional and national level, but gender gap can be observed in the propensity to collaborate at the international level. 2.2. CRIS UNS The starting point in developing CRIS UNS was creating a well-structured and comprehensive metadata set for describing scientific results, as well as researchers and institutions. Ivanovi´c et al. [23] proposed the metadata model based on the MARC 21 library standard and compatible with CERIF (Common European Research Information Format). The CERIF data model provides a very rich and well-structured set of metadata. The core of CERIF are three basic entities Person, Project and OrganisationUnit, and three result entities ResultPatent, ResultPublication and ResultProduct [17]. The structure of CERIF enabled the development of the information system in which authors are uniquely identified and connected to their results, institutions and projects. The metadata set for some entities was further enriched by introducing the MARC 21 format of bibliographic data for presenting publications and the MARC 21 format of authority data for presenting authors [41]. The described model was the basis for developing the information system CRIS UNS for tracking research activity at the University of Novi Sad [31]. Speaking of research activity, one of the main purposes for developing the CRIS UNS system was providing automated evaluation of scientific results, researchers and institutions, which has become extremely important. Paper [24] proposed an extension of CERIF by data for evaluation of published scientific results. The extension is based on the CERIF semantic layer that enables classification of entities and their relationships by different classification schemas. The rules for evaluation proposed by the academic regulatory bodies were implemented within CRIS UNS and exposed as a service for evaluation of scientific results [33]. The architecture of CRIS UNS with its rich matadata model and module for evaluation of scientific results provided the environment for developing various techniques and tools for business analysis and gathering important information used by institutional management. These tools included a module for creating periodical reports on research activity and different analyses [13]. Architecture of the CRIS UNS and adoption of international standards for presenting research data provided appropriate environment for interoperability with other systems [22] including ontology-based integration [14, 25]. Each researcher employed at our institution is obligated to have his/her CRIS UNS profile and periodically update his/her bibliographic references. Additionally, each CRIS UNS profile contains all institutionally relevant information about the researcher (academic rank, research department within the institution, year of birth, etc.) including also his/her gender. 3. GERBER methodology The GERBER methodology for gender-based analysis of intra-institutional research productivity and collaboration is based on the notion of enriched institutional co-authorship networks [38]. An enriched co-authorship network associated to an institution I is an undirected, labeled and weighted graph G = (V;E;W) with the following properties: 1. The set of nodes V corresponds to researchers employed at I and their collaborators not affiliated with I. Each node in V is labeled either as local or external – local nodes correspond to researchers from I, while external nodes represent researchers not affiliated with I. 2. The set of links E corresponds to research collaborations among researchers from V . Two researchers are connected by an undirected link if they co-authored at least one publication together (with or without other co-authors). Each link is classified either as intra-institutional (a link connecting two local nodes) or inter-institutional (a link connecting a local to an external node). 3. The function W : E ! R determines link weights which reflect the strength of research collaboration among connected researchers. The GERBER methodology uses the link weighting scheme proposed by Newman [32]. The Newman scheme takes into account the total number of authors of a paper when quantifying the strength of research collaboration between two authors. The weight of link connecting two researchers a and b is computed according to the following formula: W(a $ b) = X k2J 1 nk 􀀀 1 ; where J denotes the set of joint publications of a and b and nk is the total number of authors of publication k. The main characteristics of the scheme is that the sum of weights of all links incident to a node is equal to the total number of multi-authored publications of the corresponding researcher. 4. The network has an underlying institutional structure. This means that the set of local nodes can be partitioned into non-overlapping groups where each group corresponds to one organizational unit (research department) within the institution I. 5. A metric vector containing different types of researcher evaluation metrics (productivity, collaboration and metrics of institutional importance) is attached to each local node. The GERBER methodology extends the previously described notion with one more node attribute representing gender of researchers. This means that each local node can be classified either as male or female. Consequently we can distinguish between three types of intra-institutional links: 1. links representing collaboration between two male researchers (MM links), 2. links representing collaboration between two female researchers (FF links), and 3. links representing collaboration between male and female researchers (MF links). The GERBER methodology relies on both domain-independent and country-specific researcher productivity metrics. Three commonly used researcher productivity metrics known as normal counting, fractional (adjusted) counting, and straight counting scheme [29] are used to enrich nodes of the co-authorship network. The productivity of a researcher estimated by the normal counting scheme is equal to the total number of publications he/she (co-)authored. The straight counting scheme assigns the whole credit for a publication only to the first author, which means that the productivity of a researcher measured by this scheme is equal to the number of publications in which he/she is the first co-author (or the only author in the case of single-authored publications). The fractional counting scheme assigns credit equal to 1=n to each of n authors of a publication. Therefore, the productivity of a researcher a is equal to Pfractional (a) = X k2S 1 nk ; where S is the set of publications (co-)authored by a and nk denotes the number of authors of publication k. In our implementation of the methodology we also use a Serbian-specific researcher productivity metric known as Serbian research competency index. Namely, research publications (co-)authored by Serbian researchers are categorized according to the rule book prescribed by the Serbian Ministry of Education, Science and Technological Development. The rule book defines several categories of publication venues and each category correspond to a certain number of points. For example, papers published in the top 30% SCI ranked journals in appropriate scientific discipline are worth 8 points, papers presented at conferences are worth 1 point, etc. The competency index of a Serbian researcher is then defined as the sum of points of publications he/she (co)authored. This index is used as one of the criteria in the process of academic promotions at Serbian universities, as well as in researcher evaluation within Serbian national research projects. The degree of collaboration of a researcher with other researchers can be quantified by its degree centrality in the co-authorship network. The degree centrality of researcher a is equal to the number of links incident to a, or equivalently, to the total number of co-authors of a. Authors without collaborators (authors whose scientific output solely consists of solo-authored publications) have degree centrality equal to 0 and appear as isolated nodes in the co-authorship network. Since there are two types of nodes in the network (local and external) we can derive two additional degree centrality measures: local degree centrality – the number of local co-authors of a, and external degree centrality – the number of external co-authors of a. Clearly, degree centrality is equal to the sum of local and external degree centralities. Another local centrality measure used in our methodology is shell index associated to the k-core decomposition of networks [39]. The k-core of a graph is a maximal subgraph in which the degree centrality of each node is higher than or equal to k. The k-core can be obtained by recursively deleting all nodes whose degree centrality is smaller than k. A node has shell index k if it belongs to the k- core but not to (k + 1)􀀀core. Shell index enables us to distinguish between two types of researchers having a high degree centrality: those connected to a large number of other researchers with a low degree centrality have a low shell index, while those connected to a large number of other researchers with a high degree centrality have a high shell index. Enriched co-authorship networks are weighted graphs. Thus, we can compute weighted degree centrality for nodes. This measure reflects the strength of research collaboration of an author with other authors. The weighted degree centrality of a node is equal to the sum of weights of all links incident to the node. Additionally, the GERBER methodology includes the following variants of the weighted degree centrality measure according to different types of links in the network: local weighted degree centrality – the sum of weights of local co-authorship links incident to a researcher reflecting the strength of its intra-institutional research collaboration, external weighted degree centrality – the sum of weights of external co-authorship links incident to a researcher reflecting the strength of its inter-institutional research collaboration, intra-department weighted degree centrality – the sum of weights of intra-department co-authorship links incident to a researcher reflecting the strength of its collaboration with other researchers from the same department, and inter-department weighted degree centrality – the sum of weights of inter-department co-authorship links incident to a researcher reflecting the strength of its collaboration with institutional colleagues from other departments. Clearly, weighted degree centrality is equal to the sum of local and external weighted degree centralities, while local weighted degree centrality is equal to the sum of intra- and inter-department weighted degree centralities. Local degree centrality can be viewed as the simplest local measure of institutional importance. A researcher having a large number of local collaborators can be considered institutionally important because he/she is in the position to act as a bridge between a large number of his/her co-authors. The GERBER methodology also relies on two global node centrality metrics to quantify institutional importance of researchers: betweenness centrality [20, 28] and closeness centrality [4]. The betweenness centrality of node (author) a is the extent to which a is located on the shortest paths connecting two arbitrary selected nodes in the network. More formally, Betweenness (a) = X b;c2V;a6=b6=c (a; b; c) (b; c) ; where V denotes the set of nodes in the network, (b; c) is the number of shortest paths connecting nodes (authors) b and c, and (a; b; c) is the number of shortest paths connecting b and c that pass through a. If a large fraction of shortest paths contain a, then a can be viewed as an important node of the network in the sense that it has a vital role to the overall connectivity of the network. If the network has a clustered or community organization, then nodes with high betweenness centrality tend to be located at the intersections of communities, which means that they connect together different cohesive research groups. Betweenness centrality can be also viewed as a measure of the influence that a node has over the spread of information through the network, i.e. nodes having a high betweenness centrality are in the position to maintain and control the spread of information over the network. The closeness centrality of node a is inversely proportional to the cumulative distance between a to other nodes in the network. More formally, Closeness (a) = P 1 b2V;b6=a dab ; where V is the set of nodes in the network, and dab is the length of the shortest path connecting a and b. Nodes with high closeness centrality can be considered as socially important since they are in proximity to a large number of other nodes. In other words, nodes with high closeness centrality tend to be located in the core of the network, while nodes having low closeness centrality are located on the periphery of the network. The last category of researcher evaluation metrics present in the GERBER methodology are metrics related to characteristics of ego-networks. The ego network of node a in an undirected graph G, denoted by Ego(a), is a sub-graph of G induced by a and its nearest neighbors. The cohesiveness of ego-networks can be quantified by clustering coefficient [44]. The clustering coefficient of node a, denoted by CC(a), is the probability that two randomly selected neighbors of a are directly connected. If CC(a) = 1 then neighbors of a form the most cohesive ego-network – a clique. The lowest value of CC(a) is equal to 0 and happens when co-authors of a have never collaborated among themselves. To quantify the gender structure of ego-networks we introduce a metric called co-author gender disbalance (CGD). LetM(a) and F(a) denote the fraction of male and female collaborators of author a. Then, the co-author gender imbalance of a is defined as CGD(a) = jM(a)􀀀F(a)j. If CGD(a) = 1 then all collaborators of a have the same gender, while CGD(a) = 0 implies that a equally collaborates with male and female researchers. It is important to emphasize that global centrality metrics and ego-network metrics are computed on the reduced co-authorship network that encompasses only local researchers for two reasons: 1. The reduced co-authorship network provides an institutional boundary for the interpretation of obtained metric values, i.e. global centrality and ego-network metrics computed on the reduced co-authorship network reflect the importance of researchers within the institution and intra-institutional cohesiveness of research collaboration, respectively. 2. Co-authorship networks extracted from institutional bibliographic databases may not contain all links between external collaborators since institutional bibliographic databases do not contain full bibliographies of external collaborators, but only those publications made in cooperation with local researchers. Consequently, ego-network metrics computed on non-reduced institutional co-authorship networks may be biased towards local researchers without external collaborators. On the other hand, global centrality metrics computed on non-reduced institutional coauthorship networks cannot be used to assess the international importance of local researchers since institutional bibliographic databases do not provide a broad publication coverage in respective scientific disciplines. The GERBER methodology relies on non-parametric statistical tests applied to the sets of metric values of independent groups of nodes/links in enriched co-authorship networks in order to detect gender inequalities regarding research productivity and collaboration. The used statistical tests are the Mann-Whitney U (MWU) test [30] and the two-sample Kolmogorov-Smirnov (KS) test [18]. Let M be an arbitrary node/link metric (a metric of researcher productivity, collaboration or institutional importance in case of nodes and link weight in case of links). Let G1 and G2 denote two sets of M values for two independent groups of nodes and links. The MWU test is a rank-based test of stochastic superiority and it can be employed to test the null hypothesis that the values in G1 do not tend to be systematically smaller or greater than the values in G2. The test is based on the U statistic which is the number of times a value from G2 precedes a value from G1 in the sorted sequence of values from both groups. Under the null hypothesis U closely follows a normal distribution. The null hypothesis is rejected if the obtained p-value is smaller than 0.05, and in such cases we can conclude that there is a statistically significant difference between two groups of nodes/links regarding the aspect quantified by M. To quantify the effect size of the difference we use two probabilities of superiority [15]: 1. PS1 – the probability that a randomly selected value from G1 is strictly higher than a randomly selected value from G2, and 2. PS2 – which is the opposite probability of superiority, i.e. the probability that a randomly selected value from G2 is strictly higher than a randomly selected value from G1. PS1 +PS2 is not necessarily equal to 1: 1􀀀(PS1 +PS2) is the probability that a randomly selected value from G1 is equal to a randomly selected value from G2. The KS test checks the null hypothesis that cumulative distributions of G1 and G2 are not significantly different. The test relies on the maximal vertical distance between two empirically observed distributions (the D statistic). The null hypothesis is rejected if the obtained p-value is smaller than 0.05. The MWU and KS tests are employed to: 1. Detect gender inequalities regarding researcher productivity, collaboration and institutional importance at the institutional level and the level of research departments within institution. In other words, for each of the previously mentioned researcher evaluation metrics the tests are applied on G1 and G2 where G1 corresponds to all male researchers (resp., male researchers from a department D) and G2 to all female researchers (resp., female researchers from D) at the institutional level (resp., the level of the department D). 2. Detect gender homophily (the preference for collaboration with researchers of the same gender) in research collaboration at the institutional level, the level of intra-department collaboration and the level of inter-department collaboration. This means that the tests are used to detect statistically significant differences in link weights for all, intra-department and inter-department MM, FF and FM links, respectively. 4. GERBER tool Generally speaking, there are two options for preparing data for gender-based analysis of research productivity and collaboration: the first one is to export all relevant data to a format suitable for analysis tools, while the second one is to provide an interface for retrieving relevant data directly from a CRIS database. In this paper we opted for the first option, and discuss the second option in Section 6. GERBER is a standalone tool implemented in Java that performs gender-based analysis of data exported from the CRIS UNS system according to the methodology described in the previous section. The tool consists of three modules: Data Loader, Author Metric and Gender Analyzer. The architecture of GERBER is shown in Figure 1. We used the existing interoperability architecture of CRIS UNS to develop a module for exporting publications and authors metadata to XML documents. The Data Loader (DL) module parses two XML files that contain the data exported from the CRIS UNS information system and forms the co-authorship network of researchers appearing in the data. The first XML file contains metadata about all UNS-PMF researchers and their direct external collaborators (researchers not affiliated with UNS-PMF). Each author is described by an XML element which includes unique author identifier, author name, date of birth, institution to which the author is affiliated, organizational unit within the institution, academic rank, and gender. The second XML file contains metadata about publications which are authored by UNS-PMF researchers. Each publication is described by an XML element which consists of the following information: unique publication identifier, the complete list of author identifiers, publication year, title, publication type (journal, conference, monograph, etc.), information about publication venue and the quantitative evaluation of the publication by the rule book prescribed by the Serbian ministry of science. The extraction of co-authorship networks from CRIS-UNS data is a straightforward task since researchers present in the CRIS-UNS database are uniquely identified and consequently there are no name disambiguation problems. The DL module forms a co-authorship network in three phases. The set of nodes is formed in the first stage. The DL module iterates through the list of author XML elements and for each element creates one node in the co-authorship network. In the second phase links in the network are formed. A decentralized inverted index which maps authors to their publications is also formed in this phase. The inverted index is decentralized in the sense that a list of all publications of a researchers is directly attached to the corresponding node in the co-authorship network. The DL module iterates through the list of publication XML elements and for each publication p does the following: Connects each two authors of p by an undirected link. For each author a of publication p, adds p in the list of publications a authored. In the last phase, weights of co-authorship links are determined according the Newman weighting scheme. The decentralized inverted index is used to compute the set of joint publications for two researchers directly connected in the network. The Author Metrics (AM) module enriches nodes of the co-authorship network formed by the Data Loader module with research evaluation metrics that reflect author productivity, collaboration, institutional importance and characteristics of ego networks. Table 1 shows the complete list of metrics computed by the AM module. The productivity metrics are computed using the decentralized inverted index constructed by the DM module. Other researcher evaluation metrics are computed from the coauthorship network. Since the co-authorship network is extracted from the institutional bibliographic database it cannot be instrumented to estimate the international importance of UNS-PMF researchers. The Gender Analyzer (GA) module performs statistical comparison of male and female researchers considering researcher evaluation metrics computed by the AM module and statistical comparison of weights of different types of links. This module implements two non-parametric statistical tests: the Mann-Whitney U test and the two-sample Kolmogorov-Smirnov test. The GA module makes three types of reports: basic gender statistics considering organizational units covered by the CRIS-UNS database (different departments at our faculty), the results of the non-parametric statistical tests for the whole institution and each department within the institution, and tables that contain values of the Spearman correlation coefficient between different researcher evaluation metrics considering male and female researchers separately. 5. Results and discussion Using GERBER we performed the gender-based analysis of research productivity and collaboration at the Faculty of Sciences, University of Novi Sad (UNS-PMF). The publication and author metadata exported from the CRIS-UNS database covers 423 researchers employed at UNS-PMF and their 15097 publications written in collaboration with 5267 researchers not affiliated with UNS-PMF. The coauthorship network extracted from the exported data contains 34111 links, where 2859 links (8.38%) represent local collaborations – collaborations between UNS-PMF researchers. Only 13 nodes in the network are isolated, while other nodes belong to a giant connected component. The existence of the giant connected component in the network indicates that UNS-PMF researchers overall form a cohesive, mature research community. Table 2 shows basic gender statistics per organizational units (departments) of UNS-PMF. As can be observed, the majority of UNS-PMF researchers are female (60.76% of the total number). Female researchers are in a strong majority at the Department of Biology and Ecology and the Department of Chemistry. The smallest gender gap can be observed at the Department of Mathematics and Informatics where male and female researchers are almost equally represented. Table 2. Basic gender statistics of UNS-PMF researchers. R denote the absolute number of researchers, while M and F are percentages of male and female researchers, respectively. The results of statistical tests performed by the GA module of GERBER are summarized in Table 3. Although UNS-PMF male researchers on average have slightly higher values of all productivity metrics compared to UNS-PMF female researchers, the application of non-parametric statistical tests revealed that there are no statistically significant gender differences regarding scientific productivity. Also, we noticed that there are strong positive Spearman correlations between different productivity metrics for both genders – the lowest value of Spearman correlations for a randomly selected pair of productivity metrics is equal to 0.83. Regarding research collaboration we can see that UNS-PMF male researchers do not tend to have more both local and external collaborators than UNS-PMF female researchers, and vice versa. Collaboration metrics also exhibit strong Spearman correlations to productivity metrics (see Table 4). Moreover, the external degree centrality stronger correlates to productivity metrics compared to the local degree centrality which means that external collaborations have stronger impact to productivity for both UNS-PMF male and female researchers compared to local collaborations. From the data presented in Table 3 it can be observed that the null hypothesis of both nonparametric statistical tests were rejected for the betweenness centrality metric, but not for the closeness centrality metric. This means that there are statistically significant differences between UNS-PMF male and female researchers considering their institutional importance. Namely, UNS-PMF male researchers are not dominant in the core of the co-authorship network, but they more frequently appear as bridges that connect different, highly cohesive research groups. Statistically significant gender differences can be also observed for the clustering coefficient: ego-networks of UNS-PMF female researchers tend to be slightly more cohesive than ego-networks of UNS-PMF male researchers. This suggests that UNS-PMF female researchers tend to stimulate their unconnected collaborators to work together more often compared to UNS-PMF male researchers. Having in mind that UNS-PMF male researchers tend to have higher betweenness centrality we can conclude the following: UNS-PMF male researchers tend to be more important for the cohesion of the institution at the macro scale – they more often connect different research groups, but Table 3. The results of statistical comparison of UNS-PMF male and female researchers. hMi and hFi denote the average values of corresponding metric for male and female researchers, respectively. U is the value of the Mann-Whitney test statistic, MWU-p denotes the p-value of the MWU test, PSm and PSf are male and female probabilities of superiority, respectively. D is the value of the Kolmogorov-Smirnov test statistic and KS-p denotes the p-value of the KS test. Bold p values indicate statistically significant differences. UNS-PMF female researchers tend to be more important for the cohesion of the institution at the micro scale – they more often connect researchers from the same research group which previously have not collaborated. UNS-PMF consists of 5 departments corresponding to different scientific disciplines (see Table 2). GERBER performs non-parametric statistical tests to detect significant differences between male and female researchers within departments. Table 5 shows UNS-PMF departments and researcher metrics for which the null hypothesis of at least one of implemented non-parametric statistical test is rejected. It can be noticed that there are no statistically significant gender differences at the Department of Mathematics and Informatics and the Department of Biology and Ecology for each of considered researcher evaluation metrics. Consequently, we can conclude that there are no significant gender gaps regarding research productivity and collaboration at those two departments. Two UNS-PMF departments possessing a high level of gender imbalance are the Department of Geography and the Department of Physics. We can see that male researchers from the Department of Geography tend to have drastically higher number of external and total collaborators compared to female researchers. Additionally, male UNS-PMF geographers established significantly stronger intra-department research collaboration amongst themselves than females. This implies that males have a stronger impact to the cohesiveness of the department than females. Female UNS-PMF geographers exhibit drastically lower centrality in the co-authorship network implying that their male colleagues are more important for the cohesiveness of the department. From the results presented in Table 5 we can also observe that male researchers from the Department of Physics and female researchers from the Department of Chemistry have significantly higher co-author gender disbalance compared to their departmental colleagues of the opposite gender. Male researchers from the Department of Physics tend to have significantly higher research productivity than female researchers when research productivity is measured by the normal counting scheme. On the other hand, gender disparity in research productivity is absent when productivity is estimated by other three productivity metrics indicating that the normal counting scheme is a gender-biased research productivity metric. The normal counting scheme assigns equal credit to each author of a paper ignoring the total number of authors. Therefore, we can conclude that male UNS-PMF physicists tend to produce papers which on average have a larger number of authors than papers produced by female UNS-PMF physicists. Additionally, male UNS-PMF physicists established significantly stronger intra-institutional and inter-institutional research collaboration implying that they are more willing to collaborate with researchers from other UNS-PMF departments and researchers from other institutions compared to their departmental female colleagues. The UNS-PMF co-authorship network contains 548 links (19.17% of the total number of links) connecting male researchers (MM links), 1098 links (38.4%) connecting female researchers (FF links) and 1213 links (42.43%) representing collaborations between male and female researchers (MF links). The GERBER tool tests for statistically significant differences in the weight of MM, FF and MF links at different granularity levels. The obtained results for the UNS-PMF co-authorship network are shown in Table 6. It can be seen that statistically significant gender differences are not present at the level of the whole institution. Additionally, statistically significant differences in the weight of MM, FF and MF links are not present considering all inter-department and all inter-department links separately. However, statistically significant gender differences appear at the level of individual departments. The strength of research collaboration between male UNS-PMF physicists tends to be significantly higher than the strength of research collaboration between female UNS-PMF physicists and the strength of research collaboration between UNS-PMF physicists of opposite genders. In other UNS-PMF departments statistically significant differences in the weight of MM, FF and MF intradepartment links are absent. Therefore, we can conclude that the Department of Physics is the only UNS-PMF department which exhibits gender homophily in research collaboration. 6. Integration of analytic tools with CRIS UNS The main subject in the future work in this area is an integration of GERBER and similar analytic tools with CRIS UNS that facilitates continuous evaluation of researchers. This can be done by applying a service oriented architecture in which CRIS UNS exposes services to the analytics tools. These services will include operations for obtaining relevant data for analytic tools. Instead of dividing the analysis in two independent stages, (1) export of all relevant data to XML documents and (2) loading data from these documents, the better solution is to load data directly from the CRIS UNS database through appropriate services. Some benefits of the proposed approach are: retrieving the updated states of the entities (as CRIS UNS is in constant use by authors who enter their publications, the latest state can be obtained only by real-time access), obtaining only data that are relevant for the analysis, and omitting redundant elements that will influence memory usage, exposing the complete CRIS UNS metadata set for analysis, including the attributes that are not recognized as relevant in time of export in the current solution. As for the concrete technology and implementation of the services there are basically two solutions: implementing the web service with WSDL and SOAP and exchanging XML documents, or using the REST architectural style and exchanging either XML documents or JSON objects. Although there are some advantages of SOAP-based services in terms of tools support and type safety, REST services become very popular these days mostly due to ease of implementation based on the HTTP protocol. Table 6. Statistical comparison of MM, FF and MF link weights in the UNS-PMF co-authorship network. G1 denotes the first group, while G2 denotes the second group of links. PS1 is the probability of superiority of G1 over G2, while PS2 denotes the opposite probability of superiority. Bold p values indicate statistically significant differences. The suggested architecture is shown in Figure 2. GERBER (or some other analytic tool) accesses CRIS UNS data through the REST API providing at least the following operations: (1) retrieve all researchers for the given institution, (2) retrieve all publications for the given researcher, and (3) retrieve all publications entered after the given date. The third operation will improve the efficiency of an analytic tool such as GERBER because instead of loading all publications, it can load only those entered after the date of last access. The architecture presented in Figure 2 is applicable for any analytic tool and any research information system. In other words, if we define the complete set of operations for the REST API, and extend GERBER to load data through that API, any other research information system that implements the API can use GERBER for gender-based analysis. 7. Conclusions and future work As the main contribution of the paper we described the methodology and accompanying tool (GERBER) for gender-based analysis of intra-institutional research productivity and collaboration. The methodology is based on the application of non-parametric statistical tests to weighted co-authorship networks with an underlying institutional structure whose nodes have gender labels and associated metric vectors containing metrics reflecting productivity, collaboration and institutional importance of researchers. Additionally, the tests are conducted on different levels of granularity in order to detect gender inequalities and homophily at the institutional level and the level of research departments within institution. Using GERBER we performed gender-based analysis of our faculty (UNS-PMF) using data exported from the institutional CRIS. The obtained results showed that there are no significant gender differences at the institutional level considering productivity and collaboration of UNS-PMF researchers. On the other hand, gender differences can be observed with respect to the role researchers have in institutional cohesion – male UNS-PMF researchers tend to be more important for collaboration between different research groups, while female UNS-PMF researchers tend to be more important for collaboration within research groups. At the level of research departments situation is slightly different – there are two (out of five) UNS-PMF research departments exhibiting significant gender inequalities in research collaboration. Additionally, for one of those two departments we detected gender homophily in research collaboration. In our future work we plan to improve GERBER by including other culture-sensitive attributes such as age and academic rank in gender-based analysis of research performance. We will also investigate possibilities to enrich nodes of co-authorship networks extracted from institutional bibliographic databases with metrics reflecting the importance of researchers at the international level. To achieve this goal, GERBER has to be able to extract, retrieve or fuse field co-authorship networks having a broad coverage of individual scientific disciplines from publicly available bibliographic databases and research networking platforms [11], locate local researchers within such networks and compute centrality metrics for corresponding nodes. Additionally, our aim is to integrate GERBER into our institutional CRIS system (CRIS UNS) as an analytic service. Such integration will enable continuous gender-based analysis of researchers employed at our university and the evaluation of policies and actions conducted to reduce gender gaps in research performance. On the basis of previously mentioned integration we will also be in position to propose general methodological and technical guidelines (APIs) for culture-sensitive extensions of CRIS systems. Acknowledgments. The authors thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for support through project no. OI174023, “Intelligent techniques and their integration into wide-spectrum decision support,” and for additional support in cooperation with the Slovenian Research Agency through bilateral project no. 451-03-3095/2014-09/43, “Culture sensitive aspects in data technologies.”

Institutional logics were introduced to organisation studies by Alford and Friedland (1985) who described how contradictory practices and beliefs inherent in modern Western societies shape individuals’ actions in the political arena. The concept was popularised in their contribution, ‘Bringing Society Back In: Symbols, Practices, and Institutional Contradictions’ (Friedland and Alford 1991) where they further developed institutional logics in the context of exploring the interrelationships between individuals, organisations, and society. Meanwhile, they identified five key institutional logics - the bureaucratic state, the capitalist market, the nuclear family, democracy, and religion in the form of Christianity. Since then, there has been a significant increase in publications applying institutional logics (Reay and Jones 2016). Thornton and colleagues went on to apply institutional logics to account for the complexity of institutional changes in an institutional system (Thornton and Ocasio 1999; Thornton, Ocasio, and Lounsbury 2012). Here, institutional logics are defined as ‘the socially constructed, historical patterns of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material subsistence, organise time and space, and provide meaning to their social reality’ (Thornton and Ocasio 1999, 804). Thornton (2004) initially extended Friedland and Alford’s (1991) five logics to six institutional logics – the state, the market, the family, religion, the profession, and the corporation, in which the democracy logic proposed by Friedland and Alford was dropped out. Later Thornton, Ocasio, and Lounsbury (2012) added one more logic – community – to form seven ideal types of institutional logics (Ocasio, Thornton, and Lounsbury 2017). These seven institutional logics embody the classic formulation of logics, each of which is tightly coupled to a small number of clearly identified societal institutions. The use of institutional logics in higher education studies is a relatively new phenomenon. In a systematic literature review on the use of institutional theory in higher education, Cai and Mehari (2015) found that most higher education studies applying institutional theory refer to ‘new’ institutionalism with a focus on, for example, isomorphism and structuration processes, while more recently developed insights such as institutional entrepreneurship/work and institutional logics were rarely applied. Nevertheless, Cai and Mehari (2015) predicted a tendency towards the popularity of institutional logics in higher education research. In his review of selected higher education studies applying institutional logics, Lepori (2016) affirmed the potential of institutional logics theory for the study of higher education. He concluded that ‘logics theory could provide a more nuanced and flexible framework, which takes into account the role of (embedded) human agency and the multi-level nature of societal dynamics’ (245). The institutional logics perspective is particularly useful in higher education research because higher education is one of those ‘arenas long noted for the dominance of professionals’ (Dacin, Goodstein, and Scott 2002, 49) but increasingly represents a complex institutional system containing plural and even contesting institutional logics (Bastedo 2009; Shields and Watermeyer 2020). These changes and contestation mean that the higher education field offers a sufficient body of institutional logics research for analysis. While the concept of institutional logics has grown in popularity (including in higher education studies) due to its usefulness in helping researchers navigate complexity in studies of stable, dynamic, or emerging fields, even the originators of institutional logics themselves have identified ‘sources of confusion’ within the concept (Thornton, Ocasio, and Lounsbury 2012, 4). Suddaby (2010, 15) expresses a concern that ‘any change, however slight, is now ‘institutional’’, and we echo this concern in the area of institutional logics where any system of meaning is deemed an ‘institutional logic’, ignoring the need for ‘profound’, or field-level change (Dacin, Goodstein, and Scott 2002). Durand and Thornton (2018) point out that the ‘identification and operationalisation [of institutional logics] are not at the same level of refinement and systematic analysis as in the categories literature’ leading to delays in the development of ‘new generalisable concepts’. (650). Ocasio, Thornton, and Lounsbury (2017) express concern that the proliferation of institutional logics articles has occasioned confusion as to the conceptualisation and application of the institutional logics perspective. In doing so, they call for more research ‘on the degree of coherence of institutional logics [and how they are] differentiated from societal logics.’ (Ocasio, Thornton, and Lounsbury 2017, 511). Such a problem is even more salient in higher education research. Lepori (2016) found that most higher education studies ‘have not fully mobilised the analytical potential of the approach and the methods developed by mainstream logics studies’ (246-247). This is compounded by divergences in how institutional logics are applied in empirical analyses (Reay and Jones 2016) ranging from those that follow the ideal types identified in the classic logics literature to those that define new logics idiosyncratic to their specific research settings. As noted by Reay and Jones (2016), ‘…different authors reveal and interpret institutional logics in diverse ways, and despite the large volume of studies about logics, there is very limited discussion about how they can be identified, described, and measured’ (442). Reay and Jones (2016) identified three techniques, used by researchers in organisation studies to qualitatively capture institutional logics, namely (1) pattern deducing, (2) pattern matching, and (3) pattern inducing. These different approaches expose a tension in institutional logics studies. On the one hand, rigorous application of the seven classic institutional logics can more fully mobilise the analytical potential of institutional logics helping to ‘discern a logic and distinguish among logics, demonstrating when multiple logics are at play in a field or organisation and revealing institutional complexity’ (Reay and Jones 2016, 452). On the other hand, the theoretical development of institutional logics ‘is continuing to grow through these multiple approaches’ (Reay and Jones 2016, 452). So far, little is known about the efforts of higher education researchers to trace the development of institutional logics theory and apply it in their research to better understand the nature of higher education institutions. To fill this gap, we analyse the use of institutional logics in higher education studies by asking the following research questions: 1. What approaches to institutional logics analysis are used in higher education studies? 2. What institutional logics are identified/applied in higher education studies? 3. What challenges are evident in applying institutional logics in higher education studies? 4. How does the use of institutional logics in higher education research contribute to institutional logics theory? Methodology As our goal was to analyse literature where institutional logics are applied in higher education studies, we selected the systematic literature review as our methodology. This approach offers both transparency and rigour (Greenhalgh et al. 2004) in the attempt to answer a pre-defined research question. We followed the six steps of a systematic review process as suggested by Aguinis, Ramani, and Alabduljader (2017): 1. Determine the goal and scope of the review; 2. Determine the procedure to select journals for inclusion; 3. Calibrate source selection process through inter-coder agreement; 4. Select sources using process identified in step three; 5. Calibrate content extraction process through inter-coder agreement; and 6. Extract relevant content using multiple coders. We selected a Web of Science search due to its coverage of key databases including citation indices, conference proceedings indices, and book citation indices from Science, Social Science, and Arts and Humanities; as well as the Emerging Sources Citation Index. We acknowledge that this may miss some niche higher education journals and may bias our results towards disciplineembedded research rather than pure higher education research. We believe, however, that this method preserves the systematic nature and replicability of our review. We used a combination of keywords of ‘institutional logics’ and ‘higher education’ (included in all fields) and applied inclusion/exclusion criteria as set out in Table 1 below: The formal search, conducted on 17 September 2020, resulted in 87 publications. After a detailed review of these publications, we excluded 28 articles and selected 59 articles fitting our purpose (See PRISMA flowchart in Figure 1 and a full list of the articles in Appendix I). We qualitatively reviewed the full text of the 59 articles. Our initial coding resulted in the categorisation of data mainly in the following aspects: 1) the number of, and names given to, the logics identified; 2) whether and how these logics had been defined by the authors; 3) the contributions claimed by the authors to institutional theory; 4) the type of paper in terms of research method; 5) the approaches to identifying/applying institutional logics; 6) the disciplinary field of the publishing journal. While our coding was primarily inductive, we did compare our analysis concerning the fifth category with Reay and Jones’s (2016) three approaches of capturing institutional logics. The number of articles that investigate institutional logics in the field of higher education has increased over time – in particular over the latter half of the last decade (see Figure 2). Most of the papers are qualitative studies (48), although there are 3 conceptual papers, 6 use the quantitative method and 2 apply mixed methods. Approaches to institutional logics analysis in higher education studies Reflections on Reay and Jones’s categorisation When categorising the approaches of institutional logics analysis used in higher education literature, we initially applied the framework of Reay and Jones (2016), who identify the following three techniques used to qualitatively capture institutional logics based on their discussions with many authors of institutional logics studies: 1) Pattern deducing: ‘Gather large volume of data (primarily text), convert text to countable occurrences, and use analytic methods to reveal patterns’ (Reay and Jones 2016, 443); 2) Pattern matching: ‘Identify patterns (ideal type of logics) from extant literature and then compare data to ideal type’ (ibid.); and 3) Pattern inducing: ‘Focus on raw data using bottom-up process to identify patterns (logics) that can then be compared with extant literature’ (ibid.). Our findings indicate that although Reay and Jones’s (2016) framework is useful in observing approaches to institutional logics analysis, the articles in our review do not cleanly match their three categories. Few of the studies that we reviewed were based on the big volume data that characterises ‘pattern deducing’. The studies closest to pattern deducing were those that were based on the qualitative analysis of a reasonable number of interviews/documents. ‘Pattern matching’ is evident in organisational analysis in higher education but with two distinct application approaches. The first uses societal logics as described in classic institutional logics literature (Alford and Friedland 1985; Friedland and Alford 1991; Thornton 2004; Thornton and Ocasio 1999; Thornton, Ocasio, and Lounsbury 2012) as guiding frameworks. The second uses field-level logics identified by influential scholars in the field (e.g. Gumport 2000; Berman 2011). Regarding ‘pattern inducing’, we did find many studies that used a bottom-up process to identify ‘new’ logics in higher education studies. However, few of them went on to discuss how these identified new logics ‘can then be compared with the extant literature’ as a characteristic of ‘pattern inducing’ noted by Reay and Jones (2016). Our categorisation highlights three areas in which Reay and Jones (2016) framework could be further elaborated. Firstly, there is a difference between capturing institutional logics as discussed by Reay and Jones (2016) and higher education research using institutional logics as focused on in our study. Capturing institutional logics is about how researchers identify, describe and measure the logics (Reay and Jones 2016), whereas applying institutional logics in higher education study entails both capturing institutional logics and applying the logics for organisational analysis in empirical investigations. Secondly, we struggled to find a clear boundary between the three techniques as described by Reay & Jones. Although Reay and Jones (2016) provide a detailed comparison between the three techniques, we find overlaps between them. For instance, when describing characteristics of ‘pattern deducing’, Reay and Jones (2016) emphasise the use of analytical techniques to reveal patterns out of a big amount of data (primarily text). However, this kind of analysis could be done either in an inductive manner, which shares the same methodological ground of ‘pattern inducing’, or being guided by ideal types, which is the main feature of ‘pattern matching’. Third, while Reay and Jones (2016) see ideal-type institutional logics as societal-level logics, the ideal types of institutional logics considered by higher education researchers are at both societal-level and field-level. A new typology Our analysis reveals that the approaches to institutional logics analysis in higher education studies can be better positioned on a two-dimension typology (Figure 3). In the first dimension, we distinguish two ways of identifying institutional logics that are respectively associated with inductive and deductive reasoning. The former maps onto the ‘bottom-up process’ to identify institutional logics referred to by Reay and Jones (2016), while the latter is largely in line with ‘pattern matching’, where researchers ‘identify patterns (ideal type of logics) from extant literature’ (Reay and Jones 2016, 443). In the second dimension, we contextualise the use of the institutional logics approach based on whether the logics are identified at the societal or field level. Thornton, Ocasio, and Lounsbury (2012) posit that institutional logics concern inter-institutional systems at macro, meso and micro levels (Thornton, Ocasio, and Lounsbury 2012). Therefore, organisational analysis requires both societal and field-level logics; ‘Field-level logics are both embedded in societal-level logics and subject to field-level processes that generate distinct forms of instantiation, variation, and combination of societal logics’ (Thornton, Ocasio, and Lounsbury 2012, 148). The divisions between the approaches are not, however, always so clean cut. We did find some rare studies that use more integrated approaches. We have allocated the 59 articles to the quadrant associated with the primary approach employed. As shown in Figure 3, for each institutional logics approach, studies employ a range of research methods. While the approach of Societal-Level Induction is typically only seen in the classic institutional logics literature, the approaches discovered in our analysis of higher education studies include Societal-Level Deduction, Field-Level Deduction, and Field-Level Induction. Institutional logics identified in the different approaches and related challenges We go on to analyse which institutional logics are identified/applied in each approach and discuss the associated challenges and problems. A full list of institutional logics applied/identified in different journals is provided in Appendix II. Societal-Level induction Typical examples in the quadrant of combining societal logics and inductive reasoning are the classic literature of institutional logics, which originally identified societal logics, such as the five logics by Friedland and Alford (1991), six logics by Thornton (2004), and seven logics by Thornton, Ocasio, and Lounsbury (2012). Because our selected higher education studies focus on field level analysis, none of our reviewed studies falls into this category but we include it as a quadrant on our matrix in order to contextualise the other quadrants and offer a more complete picture of the use of institutional logics as both analytical and theoretical tool. Societal-Level deduction This quadrant includes those empirical studies that directly apply the societal institutional logics defined in classic institutional logics literature. All the eight societal logics, as ideal types of logics, proposed by the classic literature of institutional logics (i.e. the seven logics by Thornton, Ocasio, and Lounsbury 2012 combined with democracy as proposed by Friedland and Alford 1991) were mentioned in the articles in this quadrant (Figure 4). In each individual study, 2–5 logics were applied: 5 logics (2 articles), 4 logics (3 articles), 3 logics (5 articles), and 2 logics (3 articles). As shown in Figure 4, the most popular logics discussed in the literature are market, profession, and state logics. In general, studies in this quadrant more strictly follow institutional logics theory and show methodological rigour. In three studies, the authors juxtaposed new logics (managerial logic in two articles and logic of organisation in the third) with more classic societal-level logics to develop their analytical frameworks. This avoids the danger that too strict an application of the classic ideal types may exclude new logics specific to higher education. However, the need to introduce so many new logics is arguable as the majority could be replaced with existing ideal-type logics. For instance, Pietilä and Pinheiro (2020) juxtapose managerial logic with the logics of state, profession and market in their study on university career systems. Although the authors initially raised the managerial logic by citing some higher education literature, when building their analytical framework of institutional logics they based it on Goodrick and Reay (2011) and Thornton, Ocasio, and Lounsbury (2012). Based on the cited sources, we understand that the managerial logic referred to by Pietilä and Pinheiro (2020) is largely about the logic of corporation. Field-Level deduction This quadrant combines field level logics with deductive reasoning. It includes those studies that cite certain field-level institutional logics from other sources as ideal types to guide their empirical analysis. Altogether, we found 18 such logics applied in the literature in this camp (Figure 5). More detail about these ideal-type institutional logics and the sources from which the logics are cited is provided in Appendix III. Although profession, market and state logics are among the seven societal logics (Thornton, Ocasio, and Lounsbury 2012), they are treated here as a field-level logic because the authors cite field-specific sources, rather than the classic logics literature. Also, these logics are used together with other field-level logics as analytical tools. Typically, individual studies discuss two of these logics as competing with each other (e.g. professional vs. commercial, academic vs. commercial, bureaucratic vs. managerial, academic vs. market). The exceptions being those studies applying economic logics, Triple Helix logics and Era logics, each of which includes more than two sub-categories as well as a couple of studies dealing with three logics, such as market, corporate and academic logics (Louw 2019). We find three challenges arising in this quadrant. First, compared to more well-defined ideal-type logics at the societal level, agreement is lacking on ideal-type logics at the field-level. Although the institutional context of higher education is becoming increasingly complex, 18 logics in the field might be considered too many. Indeed, some logics, though with different names, share similar assumptions. For instance, Grossi, Dobija, and Strzelczyk (2020) equate managerial logic with business logic, whereas Pettersen (2015) describes a managerial logic as an instrumental logic. Excluding the Economic logics, Triple Helix logics, Sector logics and Era logics, the remaining logics can be roughly grouped according to the intrinsic similarities between them (see Table 2). This indicates that the most popular logics are academic logics (or the like), market logics (or the like) and managerial logics (or the like). Second, there are different interpretations of logics. The authors of studies in this quadrant often cite logics with the same name from a variety of sources (See Appendix III). The most consistently cited field-level ideal-type logics are ‘higher education as a social institution’ and ‘higher education as an industry’ proposed by Gumport (2000, 2003): ‘An industry logic circumscribes purposes and practices within an economic rationality, while a social institution logic enables the legitimate pursuit of a broader range of activities under the rubric of educational and democratic interests’ (Gumport 2003, 41). When applying the two logics, the authors interpret them in different ways. For instance, Paisey and Paisey (2017) consider they are represented by a corporate logic and professional logic respectively. Juusola, Kettunen, and Alajoutsijarvi (2015) interpret them as market logic and academic logic. However, corporate logic and market logics are different (Thornton, Ocasio, and Lounsbury 2012), though academic logic and professional logic are sometimes treated as exchangeable (Pettersen 2015). Finally, with only a few exceptions, such as the logics identified by Gumport (2000), the formation of most ideal-type field-level logics in higher education research is difficult to trace. In other words, we are lacking detailed explanations of why certain chosen logics can be claimed as ideal types and how they were originally developed. Field-Level induction The final quadrant, Field-Level Induction, houses those studies that inductively analyse their empirical data without initial reference to previously identified institutional logics at either societal or field level. The studies using this approach often, though not always, result in the modification and/or expansion of the range of logics at field or actor levels. Altogether, more than 30 new logics were created (See Appendix II). Many of these logics can hardly be applied to higher education at field level, but rather are idiosyncratic to the organisational settings of specific empirical studies. This approach is the most promising but, at the same time, the most problematic. It is promising because it could provide a solid basis for identifying ideal type logics in the higher education field that can be applied as analytical frameworks in empirical investigations. In so doing, it helps strengthen the approach of Field-Level Deduction. One good example is that the logics of industry and social institution in higher education, identified and elucidated by Gumport (2003) through inductive reasoning, become commonly cited ideal-type field-level logics. The approach is problematic because many of these new logics are rather freely defined and in some cases the logics identified do not strictly follow the definitions of institutional logics in the classic literature. Partially for this reason, some understandings of institutional logics focus too much on the aspect of institutional logics as tangible constructs. For instance, when institutional logics are understood as stakeholders’ beliefs (Kezar and Maxey 2014), academic disciplines (Yonezawa et al. 2020) and research excellence (Cruz-Castro, Benitez-Amado, and Sanz-Menendez 2016), there is a risk of compromising the power of institutional logics as supra-organisational ‘vocabularies of practice’ (Thornton, Ocasio, and Lounsbury 2012, 96). These problems may limit the potential of the approach to supplement Field-Level Deduction. Contribution to theoretical advancement of institutional logics by higher education studies Although our reviewed higher education studies primarily applied institutional logics as analytical tools in empirical studies, some of them also contributed to the theoretical development of institutional logics. They did so in three aspects: strengthening the core values of the theory; shedding light on blind spots within the theory; and suggesting directions for future research. Strengthening the core values of institutional logics First, the institutional logic concretises the otherwise abstract concept of ‘the institution’ by identifying a set of supra-organisational patterns that provide meaning to actions and conflicts (Thornton and Ocasio 1999, 4). Many studies reviewed in this paper indicate, either explicitly or implicitly, that their motivation in applying an institutional logics perspective was to concretely define the content and meaning of institutions in their field. Whilst the phenomena in higher education organisations are often embedded in institutional contexts, an institutional logics perspective conceptualises the abstract term of ‘context’ in a more concrete way. As such, various sets of institutional logics in the fields of higher education are identified, though higher education studies discovering ‘new’ institutional logics are also flourishing as discussed earlier. Second, the dynamic complexity of institutional logics is reflected in both horizontal and vertical dimensions in that institutional logics span horizontal and traverse vertical quadrants. On the horizontal dimension, an institutional logic perspective deals with multiple and contesting logics in institutional systems (Thornton, Ocasio, and Lounsbury 2012), rather than a more simplistic focus on dominating institutions in organisational fields (DiMaggio and Powell 1983). On the vertical dimension, an institutional logics perspective also calls for attention to analysing three levels of society, namely ‘individuals competing and negotiating, organisations in conflict and coordination, and institutions in contradiction and interdependency’ (Thornton and Ocasio 2008, 104). Such institutional complexity enables change and innovation dynamics (Thornton, Ocasio, and Lounsbury 2012). The more than 50 institutional logics mentioned in our reviewed studies clearly demonstrate logic multiplicity on the horizontal dimension. Our analysis also highlights how logics are applied/identified at societal, field and organisational levels (See Appendix II). In addition, the higher education literature further develops the power of institutional logics to explain the dynamics of institutional complexity in the context of innovation. For instance, Dudau, Kominis, and Szocs (2018) elucidate duality in innovation outcomes in the context of higher education using the insights of institutional logics. They suggest that while mixing different institutional logics may enable innovations, conflicting institutional logics, e.g. between logics of professionalism and markets, may lead to the perceived failure of innovation. Third, the institutional logics perspective explains how institutions both enable and constrain action by incorporating macro structure, local culture and human agency (Thornton, Ocasio, and Lounsbury 2012). This helps to provide a better understanding of the ‘paradox of embedded agency’ (Seo and Creed 2002): if the actions of organisational actors are constrained by taken-forgranted institutions, how and why can the actors induce institutional changes (Horton and de Araujo Wanderley 2018)? Nevertheless, its theoretical account on how mingling logics are managed via human agency remains relatively abstract. Besides demonstrating the important agency role in that actors can strategically choose institutional logics for their own benefit (Nations 2018), some higher education studies reveal micro-level mechanisms concerning strategies for balancing between competing logics or reducing tensions. For example, Narayan, Northcott, and Parker (2017) suggest the use of ‘bridging strategies’ and ‘buffering strategies’ to compromise or balance between competing logics. Mampaey and Huisman (2016) propose a typology of conflictreducing and conflict-inducing strategies for understanding universities’ responses to tensions among different institutional logics in their operating environment. Gebreiter and Nunung (2019) provide a map of different strategies for individuals to respond to conflicting institutional logics in the context of a business school. Shedding light on blind spots in institutional logics theory Cloutier and Langley (2013) pointed out four less developed areas or blind spots within current conceptualizations of institutional logics: 1) few explanations on how institutional processes play out at a micro-level, 2) little attention on struggles over conflicting logics from a legitimacy perspective, 3) a lack of consideration as to the moral aspect of institutional logics, and 4) a failure to recognise the manifestation of institutional logics in material objectives. Although the observation was made eight years ago, the higher education studies under our review largely affirmed these gaps. The contribution of higher education research primarily sheds light on the first point (as mentioned above). Relative to this, the other three areas remain less developed blind spots in organisation studies of higher education. To fill these gaps, Cloutier and Langley (2013) propose combining institutional logics with insights from other theories. Similar efforts are seen in some of our reviewed studies, such as the integration of institutional logics and imprinting theory in higher education literature (Oertel and Soll 2017; Oertel 2018). Building synergies between institutional logics and other theories would help to gain a comprehensive understanding of various organisational phenomena in the complex professional organisations that characterise the higher education field. It has already become popular to combine institutional theory (often new institutionalism) and other theories in higher education studies (Cai and Mehari 2015), though this is not a trend with respect to institutional logics analysis yet. Implications for future research Our analysis also suggests two directions for advancing the development of institutional logics theory. First, as studies applying societal-level logics and field-level logics tend to take different analysis approaches and understand institutional logics differently, there is a need to differentiate the definitions of societal-level logics and field-level logics. While the definition of institutional logics by Friedland and Alford (1991) can be understood to specifically refer to societal-level logics, the definition by Thornton and Ocasio (1999) or Thornton, Ocasio, and Lounsbury (2012) can be more flexibly applied to logics at both levels. It seems, however, that authors of our reviewed studies are not always sensitive to such distinctions. For instance, Oertel (2018, 105) applies the fieldlevel induction approach, defining institutional logics as ‘central [that] logics define means and ends and are constitutive for individuals, organisations, and society’ while citing both Friedland and Alford (1991) and Thornton, Ocasio, and Lounsbury (2012). We suggest that societal-level logics draw on institutional orders at the societal level (Friedland and Alford 1991) and have a stronger cultural component (Ocasio, Thornton, and Lounsbury 2017). Field-level logics, on the other hand, are more connected to practices, and specifically those that occur at the organisational level (Thornton, Ocasio, and Lounsbury 2012). The key differentiator for field logics is that collective identity, power and status, social classification, and attention (Thornton and Ocasio 2008) are all determined at the field level based on how things are organised or practiced, rather than the societal level based on cultural norms. This focus on practice can, however, mean that practices are sometimes labelled as logics. Both field and societal logics must, of course, maintain cultural and practice-based foundations. Thornton and Ocasio’s four criteria could, therefore, act as a checklist in order to distinguish a pure practice from a logic - identity, power, classification and attention must stem from, and underpin, the practice-based and cultural constraining/enabling of a logic. Building on this, further definition and drawing of the boundaries between field and societal logic is required in order to manage the increasing proliferation of logics at the field level. Second, the challenges regarding conceptual rigour in institutional logics analysis in higher education call for optimal methodological approaches that help realise a ‘theory-method fit’ (Gehman et al. 2017). Theory is a complex and multi-layered concept and this is exemplified in the case of institutional logics theory. Kezar (2006) distinguishes between four vertical levels of theory, namely metatheory, grand theory, middle-level theory, and low-level theory. Kezar’s key message is that there is a recursive relationship between higher and lower-level theories: higher-level theories guide and influence the theoretical development at the lower level, while lower-level theories build up to higher-level theories. Such an understanding of theory is very much in line with institutional logics, which deals with logics in societies, organisational fields and organisations (Thornton, Ocasio, and Lounsbury 2012): Logics at the field level are subject to societal-level logics, while logics at higher levels of the social structure are in turn based on the meanings produced by actors at lower levels. Thus, we propose that the four approaches to institutional logics analysis that we have outlined earlier could contribute to the building of institutional logics theory at the three theory levels (grand, middle and local) as illustrated in Figure 6. Institutional logics at the societal level can be positioned between grand level theory and middlelevel theory, whilst the field level logics sit in between middle and local levels of theory. The Societal- Level Induction approach, typically seen in the classic institutional logics literature, is the major method used to identify societal logics. Institutional logics at the organisational field level are developed by both approaches of Societal-Level Deduction and Field-Level Induction. The former examines how the societal logics are manifested at the field level; the latter discovers ‘new’ institutional logics at the field level. Once the ‘new’ field-level logics have gained legitimacy in respective research communities, they could serve as analytical tools for empirical institutional analysis in corresponding fields. Discussions as to the relationship between theory building and institutional analysis approaches imply that synergy building between inductive and deductive approaches to institutional logics analysis could better contribute to institutional research. Qualitative methods have long been used within management studies both to build theory inductively and to test theory following a deductive logic (Bansal and Corley 2012; Eisenhardt 1989). This reflects a general suggestion by the methodology literature that a research cycle that integrates deductive and inductive reasoning tends to deliver more comprehensive understandings of unknown phenomena (Newman and Benz 1998; Creswell 2003; Reay and Jones 2016). In the context of qualitative research, Langley calls this process abduction; ‘theoretical ideas, which are also out there and can be further developed’ (Gehman et al. 2017, 297). This is in line with the Gioia methodology, which recommends combining a ‘‘1st-order’’ analysis (i.e. an analysis using informant-centric terms and codes) and a ‘‘2nd-order’’ analysis (i.e. one using researcher-centric concepts, themes, and dimensions) in pursuit of qualitative rigour (Gioia, Corley, and Hamilton 2013). The inductive/ deductive combination facilitates ‘cycling between emergent data, themes, concepts, and dimensions and the relevant literature, not only to see whether what we are finding has precedents, but also whether we have discovered new concepts’ (Gioia, Corley, and Hamilton 2013, 20). Reay and Jones’ (2016) Patten Deducing (which ‘privileges analytical technics’ and is based on a large volume of data) can therefore be regarded as the first-order analysis, while Patten Inducing (which ‘privileges researcher’ and compares identified patterns with extant literature) is close to the second-order analysis. We suggest that such combined inductive/deductive approaches offer an optimum ‘theory-method fit’ in institutional logics. It offers flexibility (to reflect and respond to local logics) alongside the consistency of terminology necessary to facilitate meta-analyses and potential cross-field conversations. Concluding discussions Our analysis of the state-of-the-art application of institutional logics in higher education studies shows that the concept of institutional logics has become increasingly popular in higher education studies, especially over the past five years. Our literature analysis has revealed the usefulness of institutional logics theory in understanding universities and colleges in complex institutional environments. The higher education studies analysed in our review include research in two groups of journals: 1) those serving as primary publication outlets for higher education scholars (i.e. higher education journals and journals including higher education research as a sub-field) and 2) those including marginal research publications on higher education issues by authors from the fields of management & business as well as other areas of the social sciences (i.e. management and business journals and other social science journals). We find distinctions between institutional logics analysis in these two types of journals. Studies applying field-level ideal type logics are evenly distributed between the two kinds of journals, often focusing on tensions between two logics. Studies applying societal-level ideal type logics are more prevalent in higher education journals and tend to investigate more than two logics (up to five). This may imply that higher education researchers have heightened perceptions of the complexity of the institutional environment of higher education in which they work. Perhaps for the same reason, where ‘new’ logics are suggested in the field of higher education, this is mainly evident within the core higher education journal publications. Here, we highlight several areas that require scholarly attention to more fully exploit the power of institutional logics theory in higher education research. The first concerns conceptualising institutional logics in the context of higher education. While our reviewed articles combine to mention dozens of institutional logics, the number of logics applied/identified in each individual study does vary. So too does the rigour with which each logic is defined. This may signal a problem of ‘concept misformation’, including ‘conceptual straining’ and ‘conceptual stretching’ (Sartori 1970). These two kinds of concept misformation are associated with two potential threats to the institutional logic as a concept. One is the problem of too few logics. If eight institutional logics, as societal-level ideal types, have been identified, then there is no room for further discovery of how particular logics of specific institutions are at work. In short, there are no new questions for research. The second threat is an excessive proliferation of logics. If logics become simply a particular organisation’s engrained practices, sense of identity or sense of purpose, detached from a tight coupling with societal institutions, then an institutional logic becomes an empty concept. This divergence in the life-path of the institutional logic construct threatens its integrity and its power to increase our understanding of organisations (and the people within them) in dynamic interaction with fields. Second, the explanatory power of institutional logics has not been fully utilised in higher education studies. Between conceptual straining and conceptual stretching, is the best use of the concept: institutional logics as the tangible influence of macrological structure observed in mesoand micro-logical behaviour, routines, and artefacts (Thornton and Ocasio 2008). While our reviewed higher education studies, as a whole, do deal with institutional logics at the societal, field and organisational levels, we lack individual studies that properly elaborate how institutional logics at different levels are related and embedded. For an excellent exception see Blaschke, Frost, and Hattke (2014) which integrates macro and micro levels of analysis. Third, it is possible that some of the problems we identify may reflect the fact that some authors lean on the theoretical insights of institutional logics without a comprehensive understanding of institutional theory. Institutional logics is only one of several interrelated strands of institutional theory (Cai and Mehari 2015). For instance, to understand how various forms of institutional logics emerge, evolve, and become displaced in competing, hybrid or blended logics, other strands of institutional theory such as institutional work (e.g. Lawrence and Suddaby 2006) or institutional entrepreneurship (e.g. Battilana, Leca, and Boxenbaum 2009) can be useful. Louw (2019) offers a good example of this integration of institutional approaches. In an effort to enhance institutional analysis in higher education research, we offer our typology of approaches to institutional logics analysis as a tool for researchers that may assist in surfacing and questioning the assumptions that underpin their methodological approach. We also hope that researchers might consult our exhaustive list of institutional logics applied in higher education studies and consider the challenges we raise in the use of institutional logics in higher education research before adding new logics or loosely applying existing terminology. Our ultimate goal is that we, as a community of higher education scholars, collaboratively capture, reflect and theorise the complex institutional context in which we work and study.

There is a considerable amount of data in and about higher education systems – far more than can be analysed by official organisations and by those outside them who are formally entrusted with the task. Even the National Center for Educational Statistics in the USA, which has produced a laudably vast number of data-driven studies, cannot address all the research questions that interested parties might wish to ask. Indeed, the interests of official bodies might lie in the direction of sectoral policy-making, whereas those of institutions might be more concerned with understanding the effects of practices at institutional or subject disciplinary level and with enhancing their provision. This is where institutional researchers come into the picture (as do those whose activities could – at least in part – be given the label of ‘institutional research’, such as some in academic planning departments). Issues that transcend the individual institution can often be profitably addressed through collaborative working or through the activities of individuals who are prepared to work beyond the boundaries of their particular institutions. One such study, based on analyses of data from two institutions’ student record systems, is that of Yorke et al. (2005). Studies going beyond the boundaries of a single institution could be termed ‘supra-institutional research’. The outcomes of such studies can offer institutions potential for benchmarking their practices and/or outcomes. There is also a potential collateral benefit in that studies can avoid duplication of effort – a matter of particular significance at a time of serious economic constraint. Three examples of studies that could be labelled ‘supra-institutional research’ follow. There are plenty of others that might have been cited, such as some of the work undertaken by projects sponsored by the Higher Education Funding Council for England under its Fund for the Development of Teaching and Learning or through its Centre for Excellence in Teaching and Learning. The examples presented here are relatively low-cost studies and indicate one particular route through which knowledge, understanding and institutional practices might economically be enhanced. Example 1. The National Student Survey: an investigation in Art & Design The National Student Survey in the UK produces questionnaire data from roughly 150,000 final-year undergraduate respondents per year – a response rate of around 60%. Whilst issue can be taken with this Survey as an instrument (Yorke, 2009a) and others like it (such as the Course Experience Questionnaire used by Graduate Careers Australia), the data it generates have a robustness that few surveys manage to possess and hence complex multi-level modelling techniques can be used to analyse them in a variety of ways. Surridge’s sophisticated analyses of data from the early runs of The National Student Survey (Higher Education Funding Council for England, 2008; Surridge, 2006, 2007) have provided the higher education sector with information of considerable value to the enhancement of programme quality. Yet the information provided by such analyses may not address particular needs within the sector. For example, whilst ‘cuts’ of the response data have dealt with the relationship of responses in Creative Arts & Design to other broad subject areas, and with that of dyslexic students to students in general, the analyses produced by Surridge do not address the relationship of dyslexic students to all students within the more circumscribed grouping of subjects represented by Creative Arts & Design. (The creative arts have long been known to enrol a disproportionately high number of students suffering from dyslexia.) This is not in any way a criticism of Surridge’s estimable work, but is an indication that analyses with a specific focus may be particularly pertinent to sub-groups within the overall student population. It is unreasonable to expect official bodies such as the Higher Education Funding Council for England (which has overall responsibility for the National Student Survey) to undertake or commission every desirable analysis, since it does not have the resources necessary to do this – and, at a time when budgets are under severe pressure, the unreasonableness of any such expectation is accentuated. However, the confidentiality with which responses to the National Student Survey are treated means that the raw response data are understandably made available to a very limited number of potential analysts (who in addition need to be able to undertake complex statistical analyses). Although finely-grained analyses (such as are implied by the dyslexia example) may not be possible to researchers in general, the institutional-level data made available by the Funding Council do permit some useful analyses to be undertaken. As a broad subject area, Creative Arts & Design has consistently been rated less positively in the National Student Survey than other subject areas (especially in respect of course organisation and management), with one exception – the Survey’s scale covering assessment and feedback (where it falls into the middle of the subject areas). The group of subjects collated as Creative Arts & Design includes performance arts of various kinds; fine art; design studies; and cinematic and photographic studies. The last three of these come under the heading of Art & Design. The Group for Learning in Art & Design was concerned to ‘get under the skin’ of this apparent weakness, since Art & Design recruits strongly and is generally considered to be an area of national success (see Vaughan & Yorke, in press). The question was, at root, why an area that is apparently successful in many ways attracted less positive ratings than other subject areas. Along with the Art Design Media Subject Centre of the Higher Education Academy and the Higher Education in Art and Design Trust, the Group commissioned a study which had two prongs – an analysis of (publicly available) National Student Survey data at institutional level and a qualitative inquiry into the way a number of institutions were addressing the Survey in the light of its significance for reputation and student choice. In this article, only a segment of the first of these is discussed. The National Student Survey data for full-time students in 2009 are available for download as two large Excel files from the Unistats website (http://www.unistats.co.uk), in which the key variables are: institution; subject area (under the Joint Academic Coding System used in the UK there are three different levels of aggregation ranging from 19 to 107 subjects); whether the programme is at bachelor’s level or below; and responses to the separate National Student Survey items. The three main subjects in Art & Design are best represented by the most disaggregated grouping of subjects, but this excludes a small amount of provision that shares the same academic organisational base. In the analyses that follow, only higher education institutions that offer a significant amount of studio activity are included. Comparison of the 2007 and 2009 National Student Survey datasets (the 2008 data were not publicly available in a suitable form) show a reasonable degree of consistency in student responses in respect of the three main subjects in Art & Design, using as an index the institutional percentage of responses agreeing or strongly agreeing with the proposition of Item 22: ‘Overall, I am satisfied with the quality of the course’ (Table 1). The coefficient for Design Studies has been markedly lowered because of one institution whose responses had declined sharply in positivity between 2007 and 2009, and hence its datapoint is considerably deviant from the general trend. The enhancement-oriented questions are, first, what the institutions that show consistently strongly are doing to attract high ratings and, second, whether this is transferable. Institutions can be ordered, by specific subject, for each of the Survey’s core 22 items. This can be useful, since it can show whether an institution is bucking its general run of ratings in any particular area (analyses undertaken for a particular institution demonstrate the point). As an example, Figure 1 orders, by increasing percentage agreement with Item 22, the 66 institutions that offer Design Studies (names are replaced by rankings) and shows also the 99% confidence intervals relating to each mean. The true mean will be found 99% of the time somewhere within the range defined by the limits (but not necessarily at the computed mean): a corollary is that doubt must exist regarding a true difference between institutions if their confidence intervals overlap. It should be noted that where the number of respondents is small, the confidence limits are widely separated; where large, the limits are close together. The institution with the highest mean is scoring unambiguously better than most of the others and that with the lowest mean worse than most – in visual terms, a statistically significant difference obtains wherever confidence intervals do not overlap. Figure 1 illustrates raw data, which are probably good enough for most practical purposes. However, the raw data are influenced by the nature of the student body. If one wants to know whether an institution is performing better or worse than would be expected, it is necessary to consider the ‘residual’ score, which is generated statistically by adjusting for student demographic variables. In effect, the ‘residuals’ discount to some extent the Survey scores where the demographics are favourable to the institution and augment them where the demographics are challenging. However, residuals are scaled to a mean of zero and a standard deviation of unity, which makes interpretation difficult. Although analysis shows that there is a weak positive correlation between raw rating and residual, a number of institutions receive Survey ratings that are significantly higher than would be expected from their students’ demographics (that is, the residuals are statistically significantly different from zero: an example from the 2007 National Student Survey, for which relevant data were available, is given in Vaughan and Yorke [in press], Figure 4). Other institutions receive Survey ratings that are significantly lower than would be expected. Art & Design differs from a number of other subject areas in that it has a relatively high reliance on part-time staff who are practising artists and designers. Whilst that is a strength (in that it brings in expertise from the ‘outside world’), it creates challenges for course managers since such staff are not always available to students. Further, students have to develop a high level of self-sufficiency in that they have to negotiate access to resources of various kinds: in recent years pressure on resources has meant that students typically are not allocated a personal studio-space but have to operate according to an artists’ equivalent of ‘hot-desking’. A hint of the difference between subjects subsumed by the broad field of Creative Arts & Design and laboratory-based subjects can be seen in responses to Item 18 in the 2009 Survey (Figure 2): ‘I have been able to access specialised equipment, facilities or rooms when I needed to’. Laboratory subjects attracted on average a level of agreement some ten percentage points higher than the six subjects under the umbrella of Creative Arts & Design. The ranges of responses for each subject are, however, too wide for anything other than a tentative suggestion as to the observed difference: it is a matter that would probably be explored with greater success through focused qualitative inquiry than through more detailed survey work. A more detailed analysis than is presented here would be needed to allow consideration to be given to possible influences such as institutional status and resourcing and the entry profile of students. Why the example 1 study was worth doing Much of the rationale for undertaking this study is covered in the preceding section. A point worth adding is that reflection on what the analyses of National Student Survey data say about educational provision is based in evidence and not anecdote. Such reflection can (indeed, it probably ought to) promote consideration of aspects of ‘the student experience’ that are inadequately covered by the instrumentation – and perhaps further investigations designed to fill in the gaps. One such aspect (which is general to higher education rather than specific to Art & Design) is the relationship between what students anticipate the institutional provision to be and what it turns out to be. If the gap between expectations and experience is wide, then dissatisfaction is a potential consequence, with further consequences for Survey scores later on. Issues concerning data The unavailability of individual students’ responses to researchers other than those officially sanctioned to undertake sophisticated statistical analyses was not a problem for this study since the study was concerned with the ratings given to subject areas within institutions. Hence aggregated response data were sufficient. Some provision that would be regarded as having an academic ‘home’ in Art & Design was excluded on the grounds that the numbers were relatively small and the responses did not clearly fit within one of the three defined subjects. The omission will have had little impact on the ‘big picture’ of National Student Survey results in Art & Design. The actual analyses were straightforward to do. However a fair amount of time was spent in sorting and separating out from the two Excel tables the data relevant to Art & Design. Example 2. Honours degree classifications When results of examinations at school or higher education level are announced in the UK, there is a ritual argument about rising grades: do they signify an improvement in student performance or are they indicative of a slippage in standards? The percentage of students who gain ‘good honours degrees’ (i.e. first or upper second class honours) at bachelor’s level has risen steadily since the academic year 1994–1995, when the Higher Education Statistics Agency began to collect data systematically. This combined category is significant for students, since an ‘upper second’ or better gives access to opportunities that the lower classes of honours degree generally do not. It is also significant as a component of university ‘league tables’ (rankings) in the UK. The overall percentage of good honours degrees, however, masks some variations within the higher education sector. Some subject areas exhibit a marked rising trend (e.g. Electronic and Electrical Engineering), whereas others such as Business and Management Studies have shown hardly any increase (Figure 3). Further, there have been variations between institutional types, with the elite ‘Russell Group’ of research intensive universities showing a particularly strong rising trend between 1995 and 2002. Why the example 2 study was worth doing The issue of ‘grade inflation’ is ever-present when educational outcomes are being discussed (see, for example, Rosovsky and Hartley [2002] and Johnson [2003], who see a pernicious effect in the USA, and Adelman [2008], who is critical of general claims of grade inflation). Debate often generates more heat than light and is typically couched in terms of sector-level statistics. As well as institutional factors, disciplinary differences are often cited as being significant determinants of pedagogy and assessment and could be expected to play some part in trends in student outcomes (in the present example, degree classifications). The full analyses for the degree classification study (see Yorke, 2009b) show that institutional type and subject discipline have an influence over student awards. The term ‘grade inflation’ is often used when the word ‘rise’ might be more appropriate and certainly less pejorative. The empirical analysis prompted reflection on the possible causes of the changes over time in degree awards: not all the possible impinging factors would be expected to produce raised classifications. Identified factors (Yorke, 2009b) were as follows. Tending to raise classifications: • improvement in the quality of teaching; • greater diligence on the part of students; • learning outcomes and specific assessment criteria, which offer students the opportunity to be ‘strategic’ by targeting the expected outcomes (though perhaps at the expense of a wider engagement with the subject); • increased use of coursework in assessment, which tends to produce higher marks than do examinations (Bridges et al., 2002; Simonite, 2003); • ‘menu marking’ (Hornby, 2003), which can produce total marks that are greater than would emerge from a holistic judgement; • changes in the way classifications are determined (perhaps because of stronger results profiles from cognate institutions); and • a backwash from the publication of ‘league tables’ (rankings) of institutions. Difficult to assess the effect: • shifts in some institutions’ profiles of entering students; and • shifts in the pattern of provision because of the closure of academic departments and/or courses. Tending to lower classifications: • staff being distracted from teaching because of expectations laid on them regarding research, entrepreneurialism, administration and/or community engagement. • increasing resort by students to part-time employment (largely to fund their way through higher education). The identification of these factors provides a series of reference-points against which institutional self-evaluation and quality-related activity can be calibrated. Issues concerning data The datasets used in this study were problematic in a number of respects. The population of institutions was not stable, with institutional mergers (and the occasional demerger). The ‘noise’ thereby generated was nevertheless trivial when set against the totality of honours degree awards in England, Wales and Northern Ireland (data from Scotland were excluded because of the different structuring of the honours degree in that country). The study excluded Medicine & Dentistry and Veterinary Studies. The award of honours at bachelor’s level is rare in these subject areas. Combined studies awards were also excluded: they were by their nature heterogeneous as regards subject and in any case became of only marginal significance when the subject coding system was changed. The Open University, whose offerings were almost exclusively of combined programmes prior to the subject recoding, was – like the Scottish institutions – excluded from the analysis. The percentage of good honours degrees was based on the totality of degrees awarded to students following an honours degree programme (even if they did not achieve honours). There is a residue of non-honours degree programmes that award unclassified degrees. The unclassified degrees are supposed to be differentiated from ‘pass’ degrees, which are awarded where a student has reached the threshold for a degree award but has failed to meet the criteria for honours. Some reporting of data seems not to have differentiated the two ‘non-honours’ kinds of award. There were some obvious instances of inconsistent reporting of data against subjects (for example, where there was a gap in a run of data that normally indicated over 100 awards per year). A subject was included in the analyses if there had been at least 2000 awards per year (Nursing and Media Studies were also included – although their award numbers were below 2000 during the 1990s, both were of sufficient contemporary relevance to justify their inclusion). Subjects could only be included where a strong inference could be made that the revision to the coding system had at most a small effect. In four instances, subject data had either to be aggregated or disaggregated across the time-span of the study in order to reflect differences in the coding structure (for example, Business and Management were originally combined in the coding system, but separated in the revised version: for the purposes of the study, the data were combined throughout as Business & Management). The changes in subject coding that were introduced for the academic year 2002–2003 are very likely to have had the strongest influence on the analyses. Whilst, in general, the codings were consistent with those previously in operation, in some instances the change in structure of the coding system made continuity of subject data problematic, and a number of subjects had to be excluded on that count. Prior to 2002, awards were coded against 18 broad subject areas plus the ‘Combined’ category. The Combined category swept up all awards that could not be related to a single subject area. As noted against Figure 3, awards after 2002 from combined programmes were roughly subdivided between the relevant subjects. Examination of the numbers of awards suggests that for some subjects the change made little difference, whereas for other subjects the number of awards showed a marked discontinuity with those that had gone before. The data for the academic year 2002–2003 are likely to be less robust than those for succeeding years because of the need for institutions to come to terms with the changes to the coding of subjects. Inspection of the datasets suggests that a few institutions had not fully come to terms with the new requirements. A very considerable amount of work had to be done to prepare the data in a form suitable for analysis. The Higher Education Statistics Agency provided a matrix of institutions x subjects for each level of the honours classification. The matrices were not identical in size, since rows and columns were omitted where there was nothing to report (rather than were simply left blank). This meant that the matrices first had to be brought to a common size, so that rows and columns aligned. For analyses by institutional type (not reported here), the relevant institutions had to be extracted from the overall matrices and their data aligned. This exercise required a judgement to be made regarding to how to categorise some institutions, since a number of colleges became universities during the period in question. However this was done, there would inevitably be anomalies, such as the multi-disciplinary college (categorised under ‘Colleges’) that was not markedly different from a similar institution that had gained university status (categorised as a ‘new University’). In ‘big picture’ terms, the chosen categorisation of institutions probably had only a marginal effect on the outcomes of the analyses – an analogy can be made with stock market indexes, which evolve over time as institutions drop out because they no longer fit the criteria for admission, to be replaced by others that do. Example 3. Institutional assessment regulations The ‘Burgess Group’ was established in the UK to review a number of issues relating to the recording of student achievement: value added; degree classifications; and credit systems. With evolving membership it was tasked with scoping out the issues and then with proposing resolutions of the issues that had been considered. The Group produced a sequence of three publications relevant to this article: Universities UK and Standing Conference of Principals (2004, 2005) and Universities UK and GuildHE (2007). The Student Assessment and Classification Working Group (an informal group of academics and administrators, set up in 1994, who share an interest in assessment matters) was commissioned by the Higher Education Academy to undertake a survey of institutional assessment regulations to inform the Burgess Group’s review of the ways in which student achievement in the UK was recorded. The Student Assessment and Classification Working Group had, over more than a decade, undertaken research focusing on the honours degree classification and had considered the impact that assessment regulations could have. The survey covered 35 institutions of varying types, this sample being reasonably representative of institutional diversity in the UK. The institutional assessment regulations were examined with reference to a set of issues that this Group had previously identified as having significance for the classification process and, where necessary, clarification of points of detail was undertaken with institutional correspondents. The outcomes of this analysis showed that institutions had their own particular sets of regulations and between them there existed considerable variation. This variation could be seen, for example, in: • the grading system (‘percentage’ marks or other kinds of scale); • the ways in which grades were combined to arrive at an overall classification (including the relative weightings given to work in the penultimate and final years of full-time study); and • the extent to which allowances could be made in respect of failures, personal misfortune, and the like. A detailed account of this study is given by Yorke et al. (2008). If the results from the sample can be generalised across the higher education system in the UK (a not unreasonable assumption), it is very difficult indeed for any interested party to reach a definitive judgement as to the relative merits of any given class of honours degree (in, say, Sociology) from Institutions X and Y. Clearly the very good and very weak students’ awards are unlikely to be susceptible to significant amendment were they to be run through the classification process of a different institution. It is the students whose results fall close to a classification boundary (in a majority of institutions, the boundaries are at an average mark of 40, 50, 60 and 70%) who might benefit or suffer from changes in assessment regulations – and the potential for a different classification is greatest at the boundary between upper and lower second class honours, which is (as noted earlier) a significant determinant of career chances. Why the example 3 study was worth doing There has been a longstanding debate in the UK about the extent to which honours degree classifications are comparable – the parliamentary inquiry into Students and universities revealed the difficulty that a number of vice-chancellors found when pressed by the members of parliament who were conducting it (see Innovation, Universities, Science and Skills Committee, 2009, passim). Issues that come to the fore include the level of academic demand and the methods used to assess the extent to which the academic demand is met. Relatively little attention has been given to the regulations that are applied in respect of assessment. Early work by the Student Assessment and Classification Working Group showed that honours degree classifications could vary if students’ marks were run through different institutions’ methods for determining classifications (Woolf & Turner, 1997). Subsequently, Yorke et al. (2004) showed that the extent to which the weakest result(s) of a student could be dropped from the determination of the classification (‘dropping’ weak results is a feature of some classification systems) could have, as one would expect, the effect of raising the (notional) award. Changing the relative weightings of penultimate year and final year results could alter the notional classification, but in this case the shift could be upwards or downwards. When the Burgess Group was reviewing the honours degree classification (and other matters), it needed information about the extent of variation in the higher education system regarding the determination of honours degree classifications – hence the commissioning of the study. The considerable variation disclosed by this study implicitly asks whether there should be some convergence in assessment regulations across the higher education sector in order to mitigate one of the difficulties associated with the issue of comparability of academic standards. A first step would be for institutions, on a collective basis, to discuss the various differences of view manifested in assessment regulations, with the intention of clarifying and testing out the underlying assumptions of assessment regulations. That step has yet to be taken. Issues concerning data In this particular case, there were few problems with the quality of the data. Institutional documentation on assessment regulations was extensive, though there was considerable variation in the way that this was laid out. There were some points at which institutional documentation was unclear and the cross-checking with the institution concerned was essential to obtain an understanding of exactly what was being meant (even the most tightly-drawn regulations seem to embed some tacit understandings). The biggest challenge was to lay out, against the grid of previously-identified issues, the relevant details from the 35 sets of regulations. Some of the data was in a form that did not lend itself easily to the kind of categorisation adopted (the same would have been true for alternative approaches to analysis). It was generally not the case that institutions provided, in their documentation, a rationale for the regulations that they had adopted. When a few were asked about this, it seemed as though the regulations had evolved over time, save where there was a major shift in curricular structure – of which the widespread adoption of modular schemes in the UK during the 1980s and 1990s was the most prominent. Commentary A spectrum of activities can be subsumed under the heading of ‘supra-institutional research’, ranging from the large-scale and heavily resourced studies (such as that of Bowen, Chingos, & McPherson, 2009) to relatively informal Internet inquiries of colleagues who share electronic membership of a group such as ‘Improving Student Learning’. If the ends of this spectrum can be characterised as ‘macro-level’ and ‘micro-level’, then the examples presented in this article are probably best seen as ‘meso-level’, in that they involved quite significant amounts of work. It is not the purpose of this article to deal in depth with the individual studies; readers interested in them can consult the original references if they wish for greater detail. The examples illustrate that studies whose usefulness extends beyond a single institution can be undertaken at relatively small cost (none of the examples cost more than £10,000). At a time of considerable economic stringency, the benefit/cost ratio of such activities has a particular attraction. Volkwein (1999) discussed institutional research in terms of a fourfold typology – the potential audience (internal or external to the institution) and focus (institutional and administrative or academic and professional). The meso-level examples in this article are relevant to varied constituencies in Volkwein’s typology. The limitations on the magnitude of meso-level studies imply that the full rigour of educational research (as some might expect of the ‘external x academic and professional’ cell of the typology) might not be attainable. The studies might need to pass a slightly lower threshold – that of being ‘good enough’ for the purposes for which they are intended. They still have to be good studies. ‘Good enough’ does not imply that ‘anything goes’; rather, it makes it particularly important that the researchers openly acknowledge the limits inherent in their work. There is considerable potential, at the meso-level, for useful studies to be undertaken, provided that ‘buy-in’ from a number of institutions can be achieved. Institutions might be willing to contribute very modest sums of money towards studies that address issues of common interest. The totality of contributions could then allow to be achieved much more than could be done on a single-institution basis: there is a potential economy of scale. Institutions might also need to contribute not only in cash terms but also ‘in kind’ – for example, by preparing datasets to a common format or by organising access to potential respondents to questionnaires. ‘The deal’ would be that they would get outcomes of the study that go beyond the individual institution, against which their own data can be compared in a kind of benchmarking exercise. This general approach proved essential in a study of part-time students’ experiences in post-1992 universities in the UK (Yorke & Longden et al., 2008) since, despite the importance of part-time higher education in governmental policy, funding to investigate this little explored topic was not forthcoming. In the end, a residue of funding from an earlier project was a sufficient resource to support the activities of the project team and participating institutions contributed their team members’ time and engagement. Without such an approach, nothing of any magnitude could have been achieved. There are issues of trust to be factored into the arrangements. For example, the complete dataset may have to be restricted to the lead researcher(s) since data from individual institutions may be deemed too sensitive for general sharing within the project team and some freshly-gathered data may – if they are to be returned to the originating institution for further analysis – need to have potential identifiers removed in order to respect respondent anonymity. ‘Supra-institutional research’ is not new: there are plenty of examples down the years. Indeed, unlike Monsieur Jourdain in Molière’s Le bourgeois gentilhomme, some have spoken the prose of supra-institutional research for a long time, and have realised that they were doing so. However, in respect of the meso-level that is the focus of this article, it may be necessary for the speaking to be a little louder.

Over the past four decades, research informed by institutional theory has emerged as one of the main strands of sociological thought in the accounting research literature. This strand of thought has been especially influential in management accounting (MA) research where it has been deployed to a range of topics, such as budgeting (Covaleski and Dirsmith 1983; Fernandez-Revuelta Perez and Robson 1999; Kaufman and Covaleski 2019), costing (Ansari and Euske 1987; Modell 2002; Krishnan and Yetman 2011), performance measurement (Brignall and Modell 2000; Carlsson-Wall et al. 2016; Modell 2019) and the changing professional roles of management accountants (Burns and Baldvinsdottir 2005; Goretzki et al. 2013; Horton and Wanderley 2018), to name but a few. However, in recent years, some contributors to this body of research have started to raise concerns about the alleged tendencies of accounting scholars to apply institutional theory in an uncritical and overly eclectic manner in their efforts to advance novel insights into the roles of accounting in organizations and society (Modell 2015, 2022;Modell et al. 2017). These concerns echo those emerging in the broader organization studies literature. Even though the many variants of institutional theory, which collectively make up the larger research program of new (or neo-) institutional sociology, have been celebrated as one of the main strands of thought in contemporary organizational research (Lounsbury and Beckman 2015; Greenwood et al. 2017), critics have argued that this program has reached a point of intellectual exhaustion where more attention is being paid to repackaging institutional arguments than providing substantially novel insights into organizational behavior (Alvesson and Spicer 2019; Alvesson et al. 2019; Reed and Burrell 2019; Aksom and Tymchenko 2020; Aksom et al. 2020). Invoking a Lakatosian perspective on scientific progress (Lakatos 1970), some critics have gone as far as arguing that institutional theory now constitutes a degenerative research program that has largely ceased to make progressive contributions to organizational research (Reed and Burrell 2019; Aksom and Tymchenko 2020). Such degenerative tendencies may be amplified when theories are transposed to other fields of research such as accounting. For instance, concerns have been raised that the extensive borrowing of theories from cognate fields of research in contemporary accounting research does not necessarily help in advancing our knowledge of MA as an organizational phenomenon (Malmi and Granlund 2009; Baldvinsdottir et al. 2010). Concerns have also been raised that accounting scholars do not always make progressive contributions back to the fields from which theories are borrowed based on insights that are somehow unique to their field of research (Lukka and Vinnari 2014). Applying the Lakatosian model of scientific progress to institutional research on MA, I examine the extent to which this body of research has contributed to degenerative and progressive tendencies in the accounting literature as well as in the larger institutional research program in organization studies. Based on a systematic review of MA research drawing on different variants of institutional theory, I offer a more nuanced assessment of the achievements of this research program than what has emerged from the recent, critical debates around the use of institutional theory in accounting research and organization studies. I show that this body of research has made a number of progressive extensions to our understanding of how accounting practices are (de-)institutionalized while, to a lesser extent, offering similar extensions to the larger institutional research program. However, there is also evidence of many MA scholars using institutional theory in ways that produce degenerative tendencies. These degenerative tendencies are, in large part, due to the persistent proclivity of researchers to bracket the role of either human agency or preexisting institutions and, thereby, place overly one-sided emphasis on one or the other in explaining the process of (de-)institutionalization. These tendencies may, at worst, cause functionalist assumptions to be smuggled back into institutional analyses in ways that threaten the hard core of the institutional research program. I discuss how these tendencies can be rectified in future research. In doing so, I urge accounting scholars to continue to advance analytical approaches that place relatively balanced emphasis on agency and extant institutions. Although such a perspective can be applied to a range of MA topics, I stress the need to combine it with a focus on the constitutive role of accounting in the process of (de-)institutionalization. This is one of the few areas where MA scholars have utilized insights that are somewhat unique to their field of research to offer progressive extensions to the larger institutional research program. Such research also constitutes a powerful bulwark against any tendencies to smuggle functionalist assumptions back into institutional analyses as it compels researchers to remain faithful to the social constructivist underpinnings of institutional theory. I discuss how such research may be further developed. The remainder of the paper proceeds as follows. I start by outlining the Lakatosian model of scientific progress and considering if the larger research program that makes up institutional theory is degenerating or progressing. I then extend this analysis of degenerative and progressive tendencies to the field of MA before concluding the paper with a discussion of key findings and the implications for future research. 2. Institutional theory as a Lakatosian research program The Lakatosian model of scientific progress The Lakatosian model of scientific progress builds on an evolutionary view of theory development and is applicable to anybody of scholarship that has a clearly articulated explanatory purpose. 1 In contrast to Kuhnian notions of scientific progress following radical paradigm shifts (Kuhn 1962) and Popperian notions of falsification as the key criterion for refuting theories and replacing them with new ones (Popper 1959), Lakatos (1970) suggested that theories do not necessarily replace each other through discontinuous breaks. Instead, Lakatos (1970) proposed a view of scientific progress as a process whereby a number of related theories2 that make up a larger research program, emerge and continue to coexist over extended periods of time. Such programs are progressive insofar as each new theory offers improved explanations of a particular phenomenon without contradicting existing theories within the same program. The search for such improved explanations is often triggered by empirically observable anomalies, which existing theories have not yet explained but which new theories are able to expound. For a research program to be genuinely progressive each new theory needs to be both theoretically and empirically progressive. Theories are theoretically progressive insofar as they have some additional conceptual content that goes beyond their predecessors and enables researchers to explain perceived anomalies. To be empirically progressive, new theories also have to be corroborated when confronted with empirical data. Unless both of these criteria are met, Lakatos (1970) calls research programs degenerative. Degenerative research programs often emerge as researchers try to build, but fail to maintain, a protective belt of additional contents, or complementary theories, around the hard core that gives research programs their distinct identity. The hard core of a research program is made up of the fundamental assumptions about the phenomenon to be explained that are shared across the theories that constitute the program and that cannot be refuted without the entire program being abandoned. By contrast, the protective belt of complementary theories can be continuously adjusted as researchers seek to extend their program. However, unless such extensions enhance the theoretical content of the research program as a whole and receive empirical support, the program will degenerate over time. Such degenerative tendencies might, in turn, threaten the hard core of the program and possibly cause it to be replaced by competing programs (Lakatos 1970). Having outlined the Lakatosian model of scientific progress, it is important to address at least two challenges to this model. One challenge is whether Lakatos’s (1970) evolutionary perspective on science, which assumes that scientific progress is of a fundamentally incremental nature, builds on a natural science view of scientific work that is not applicable to the social sciences. Although there are mixed views as to whether Lakatos considered his model relevant to the social sciences, there is ample evidence to suggest that this was indeed the case (Elman and Elman 2002). Indeed, the Lakatosian conception of scientific progress has had considerable influence on discussions of what constitutes a theoretical contribution in organization studies (Whetten 1989; Kilduff 2006; Corley and Gioia 2011), although organizational scholars also recognize that scientific progress may entail radical shifts in scientific knowledge claims (Suddaby et al. 2011; Corley and Gioia 2011).3 A similar, incremental view of scientific progress is commonplace in contemporary accounting research and can be found in not only positivist research (Mouck 1990; Boland and Gordon 1992), but also inter-disciplinary accounting scholarship informed by a wide range of theoretical perspectives (Modell et al. 2017; Richardson 2018). The Lakatosian conception of scientific progress also builds on a view of scientific work as an unfolding social endeavor that is governed by inter-subjective rather than strictly objective criteria. Hence, what counts as emerging, progressive extensions of a research program at any given time needs to be determined with reference to what adherents to the program see as the state of the art of the program at that point in time (Gholson and Barker 1985). However, for this view of what constitutes scientific progress not to give way to a hyper-relativist, anything-goes position (Feyerabend 1975), there needs to be a broadly based consensus about what constitutes the hard core of a particular research program among the scholars who subscribe to such programs. This brings us to the second challenge to Lakatos’s (1970) view of how research programs are constituted, which is notably associated with Laudan’s (1977) critique of the assumption that such programs have a stable core. As an alternative to this assumption, Laudan (1977) argued that scientific work falls into broader research traditions that share certain ontological commitments but where these commitments can change and vary over time. Although this critique does not invalidate the notion that the hard core of some research programs is relatively stable (Elman and Elman 2002), it underlines the need to delineate carefully the assumptions around which there is a broadly based consensus and that need to remain unchanged for the core to be protected from degenerative tendencies. These assumptions also need to be distinguished clearly from other features of the program that may change and vary somewhat as it evolves and that may give rise to degenerative tendencies that do not constitute an immediate threat to its hard core. This is especially important when analyzing relatively heterogeneous research programs, such as that of institutional theory, and determining whether such programs are degenerating or progressing. Is institutional theory degenerating or progressing? Applying the Lakatosian model of scientific progress to the many variants of institutional theory that have evolved over time, we first need to specify what may be seen as the hard core of this larger research program and then determine whether this core has remained reasonably stable over time. Starting with the seminal works of Meyer and Rowan (1977) and DiMaggio and Powell (1983), what gradually came to be known as new (and, somewhat later, neo-) institutional sociology emerged as a distinct break with older, functionalist theories of organizations, such as contingency theory. Growing out of the social constructivism of Berger and Luckmann (1967) and a practice-based view of organizations that was heavily inspired by Bourdieu (1977), institutional theory is unified by a view of organizations as situated in institutional fields, or recognized areas of social life, and an emphasis on the institutionalization and de-institutionalization of organizational practices as the key phenomenon to be explained (Greenwood et al. 2017).4 Institutionalization refers to the processes through which organizational practices come to be taken for granted and imbued with legitimacy, while de-institutionalization is the process through which such practices are challenged and change. In contrast to functionalist approaches, which see organizational design choices as an outcome of instrumental quests for fit between organizations and their environments in an attempt to achieve superior performance, institutional theorists see the need for social legitimacy as key to a conception of organizations as socially constructed entities. These assumptions, which are widely shared among institutional theorists, imply a rejection of rational actor models of organizations in favor of a view of human agency as conditioned by the institutions that evolve in institutional fields (Greenwood et al. 2017; Zietsma et al. 2017). Without a basic recognition of organizations as socially constructed entities, whose persistence is ultimately dependent on their ability to achieve and maintain social legitimacy, and a notion of agency as a socially conditioned phenomenon, institutional theory arguably loses its distinctiveness from functionalist approaches. Research that poses as institutional without recognizing these assumptions and translating them into the conceptualization of (de-)institutionalization is likely to threaten the hard core of the institutional research program. Around the hard core of a view of organizations as socially constructed entities, whose survival is determined by their ability to achieve and maintain social legitimacy, and human agency as a socially conditioned phenomenon, institutional theorists have sought to build a protective belt of additional, though quite varied, contents in response to perceived anomalies. As explicated below, all of these variants of institutional theory have a commitment to this hard core, although the relative emphasis on human agency and preexisting institutions as forces explaining the process of (de-)institutionalization varies somewhat. As we shall see, these variations are, in large part, due to the bracketing of either agency or institutions in institutional analyses and the efforts that have been made to address this problem within the institutional research program. Such bracketing can give rise to degenerative tendencies insofar as it leads researchers to advance explanations of (de-)institutionalization that do not add any theoretical content to prior advances and, therefore, fail to improve on extant explanations. However, it does not necessarily constitute an immediate threat to the hard core of the institutional research program. Such threats are only imminent where researchers wittingly or unwittingly challenge the assumptions that underpin this core, as may be the case where institutional arguments are subjected to critical tests or researchers smuggle functionalist assumptions back into institutional analyses. Early advances in institutional theory arguably subscribed to a relatively deterministic view of institutionalization and posited that the quest for social legitimacy leads organizations within the same institutional field to become relatively similar as they adopt and reproduce institutionalized practices (Meyer and Rowan 1977; DiMaggio and Powell 1983; Tolbert and Zucker 1996). Such tendencies toward institutional isomorphism were seen as a nearly inevitable consequence of the process of institutionalization and a powerful source of stability in institutional fields (DiMaggio and Powell 1983; Scott 1995). Also, organizational practices that are adopted for the sake of legitimacy-seeking were often seen as being in conflict with the technical, or functional, requirements of organizations and thus decoupled from operations in order to prevent disruptive conflicts that may threaten institutionalized practices (Meyer and Rowan 1977). However, the emphasis on homogeneity and stability soon came under fire for bracketing the role of vested interests, power and deliberate, strategic agency in institutional processes and thereby failing to explain change and the possibilities of resisting institutional pressures. Specifically, DiMaggio (1988) advanced the notion of institutional entrepreneurship as a corrective to this anomaly. Institutional entrepreneurs are generally seen as actors who initiate changes that diverge from institutionalized practices and who devise strategies for transforming organizations and institutional fields (Hardy and Maguire 2008; Battilana et al. 2009). This renewed attention to institutional change was subsequently followed by conceptual advances that sought to expound the broader range of strategic responses beyond decoupling that enable organizations to resist institutional pressures (Oliver 1991) and how processes of de-institutionalization are initiated (Oliver 1992). Although these early attempts to explain institutional change and resistance opened up new vistas for institutional theorists, they were soon criticized for placing too much emphasis on rational, strategic agency exercised by individual change agents and, thereby, contributing to a development that might cause the institutional research program to degenerate (Holm 1995; Hirsch and Lounsbury 1997; Dacin et al. 2002; Seo and Creed 2002). This is not to say that these advances ignored the assumptions underpinning the hard core of the institutional research program. For instance, both DiMaggio (1988) and Oliver 1992 recognized the role of preexisting institutions as a socially constructed, constraining force that conditions the possibilities of human agency. However, especially Oliver (1991) combined institutional theory with strategic choice theory (Child 1972) and later juxtaposed institutional and functional pressures as distinct and potentially competing antecedents of de-institutionalization (Oliver 1992). As such, she can be said to have introduced an element of voluntarism in institutional theory that may invite researchers to bracket the influence of extant institutions and, at worst, smuggle functionalist assumptions that foreground rational, interest-driven quests for superior performance rather than broader, socially constructed concerns with legitimacy into institutional analyses. If borne out empirically, this view would effectively threaten the hard core of the institutional research program and possibly accelerate its demise. However, throughout the evolution of this program, there have been few attempts to subject the assumptions that underpin this core to critical tests (Davis 2010, 2015) and institutional theorists have, by and large, steered clear of functionalist concerns with organizational performance (David and Bitektine 2009; Lockett et al. 2015). Although notions of institutional entrepreneurship and strategic agency have not completely disappeared from institutional analyses, they were soon superseded by advances that foregrounded the social conditioning of human agency without reverting to deterministic conceptions of (de-)institutionalization such as those associated with research on institutional isomorphism (Holm 1995; Greenwood and Hinings 1996; Seo and Creed 2002; Battilana 2006). Also, insofar as notions of institutional entrepreneurship are still in use, such entrepreneurship is now explicitly framed as a socially conditioned phenomenon (Hardy and Maguire 2008; Battilana et al. 2009). A key premise uniting these advances is that any agency evolving in institutional fields is always embedded in preexisting institutions and that such embeddedness conditions, albeit not determines, the efforts of human beings to reproduce as well as transform institutions. Institutional embeddedness operates at multiple levels of analysis and can be traced to diverse mechanisms, ranging from the social structures that underpin institutional fields to the structuring of individual organizations within such fields and the institutionalized identities that shape the action repertoires of individual human beings. These mechanisms shape the world view of actors in institutional fields and not only constrain but also enable the agency involved in the (de-)institutionalization of organizational practices. However, the increasing emphasis on embedded agency gave rise to a new anomaly that is widely referred to as the paradox of embedded agency (Seo and Creed 2002; Battilana and D’Aunno 2009). This paradox pivots on the question of how agents who are embedded in extant institutions can initiate change in the very same institutions. To address this conundrum, which is especially germane for explaining processes of de-institutionalization, institutional theorists began to trace the emergence of institutional change to institutional contradictions, such as the existence of incompatible institutions, which compel agents to deliberate on and, to a degree, disembed themselves from extant institutions (Clemens and Cook 1999; Seo and Creed 2002; Greenwood and Suddaby 2006). Such conceptualizations refined the view of human agents as embedded in extant institutions but, at the same time, capable of initiating intentional change in such institutions without rejecting the possibility that some institutional fields are still relatively homogeneous and stable. This can be seen as a progressive extension of the institutional research program that avoids the earlier tendencies to bracket either the role of agency or that of extant institutions in institutional processes and that maintains a balanced emphasis on both in explaining the process of (de-)institutionalization. One of the main strands of institutional thought that has at least conceptually sought to avoid such bracketing is that pivoting on the notion of institutional work. Originating in the work of Lawrence and Suddaby (2006), this approach has evolved into one of the two strands of thought that currently dominate the institutional research program (Zilber 2013; Hampel et al. 2017). Lawrence and Suddaby (2006, 215) defined institutional work as “the purposive action of individuals and organizations aimed at creating, maintaining and disrupting institutions,” but insisted that such work is always conditioned by extant institutions. Also, rather than focusing on individual change agents, research on institutional work pays more attention to the distributed nature of agency in institutional fields and the complex processes through which individual agency is transformed into collective agency (Lawrence and Suddaby 2006; Hampel et al. 2017). To capture such processes, propagators of the institutional work perspective have developed extensive taxonomies that allow researchers to disaggregate the agency involved in the (de-)institutionalization of organizational practices in considerable detail and examine the interplay between different types of agency (Lawrence and Suddaby 2006; Perkmann and Spicer 2008). At one level, research on institutional work can be seen as a progressive extension of the institutional research program that furthers our understanding of embedded agency. Moving beyond the relatively rare occurrence of institutional contradictions as a source of large-scale, institutional change, research on institutional work has arguably refined our understanding of the more mundane processes through which (de-)institutionalization unfolds without reverting to notions of human agency being exercised by relatively unconstrained actors (Battilana and D’Aunno 2009; Hampel et al. 2017). Research on institutional work has also been credited with bringing notions of power and politics to the fore in institutional analyses without submitting to naïve, instrumentalist notions of political processes as manageable by omnipotent change agents (Lawrence et al. 2009; Hampel et al. 2017). However, critics of the institutional work perspective have argued that, despite its conceptual efforts to avoid bracketing the influence of extant institutions, empirical inquiries have become too preoccupied with unpacking the agency involved in the process of (de-)institutionalization at the expense of systematic attempts to explain how such institutions influence this process (Khagan and Lounsbury 2011; Ocasio et al. 2017; Modell 2022). This blurs the distinction between research on institutional work and studies of institutional entrepreneurship and strategic agency, from which it initially sought to distance itself, and can reinforce the degenerative tendencies associated with a lack of attention to embedded agency. Even though this does not necessarily imply a return to functionalist accounts of organizations and a threat to the hard core of the institutional research program, it can lead researchers to overemphasize the intentional, interest-driven nature of the agency that constitutes institutional work. Besides research on institutional work, the other strand of thought that currently dominates the institutional research program and that has sought to avoid bracketing the role of either agency or extant institutions in the process of (de-)institutionalization is that pivoting on the notion of institutional logics (Zilber 2013; Ocasio et al. 2017). Institutional logics are generally conceived of as symbolic and material templates for organizing that originate in broader societal orders, such as corporations, markets and the nation state, and that condition any notions of agency evolving in institutional fields (Thornton et al. 2012). Originating in the work of Friedland and Alford (1991), much research on institutional logics has been motivated by the alleged tendency for research on institutional isomorphism to over-emphasize the homogeneity of institutional fields and failing to explain how variations in organizational practices emerge in such fields (Lounsbury 2008; Thornton et al. 2012). Addressing this anomaly, research on institutional logics has tended to deemphasize, albeit not reject, notions of institutional isomorphism in favor of notions of institutional complexity (Greenwood et al. 2011; Thornton et al. 2012). Institutional complexity is typically seen as a product of multiple, competing logics that compel organizations within the same field to pursue either different strategies that follow distinct logics or internalize multiple logics into intricate organizational arrangements that cause such logics to be separated from each other or combined into hybrid organizational practices (Thornton et al. 2012: Ocasio et al. 2017). Another important insight from the institutional logics perspective is that logics shape not only what is seen as legitimate, but also functional and rational, in institutional fields and that even seemingly strategic forms of agency need to be conceived of as embedded in extant institutions (Lounsbury 2008; Thornton et al. 2012). In contrast to the relatively sharp distinction between functionalist and institutional explanations of organizational behavior permeating research on institutional isomorphism, this leads to a view of rationality as a socially constructed phenomenon that is not possible to disentangle from the logics in which it is embedded. To assess whether research on institutional logics represents a degenerative or progressive extension of the institutional research program, we first need to answer the question of whether its explanations of how variations in institutional fields occur are substantively novel or if they were, in fact, anticipated by older variants of institutional theory. On this point, the propagators of the institutional logics perspective are somewhat ambivalent. In his introduction of this perspective to an accounting audience, Lounsbury (2008, 353) observed that, to Meyer and Rowan (1977), isomorphism was “not a mindless, structurally determined process, but an effortful accomplishment,” that recognizes that “multiple forms of rationality may exist, providing a foundation for the explanation of organizational variety.” Lounsbury (2008) also intimated that the very notion of decoupling, as articulated by Meyer and Rowan (1977), implies that what may appear to be highly isomorphic tendencies at the overriding field level may actually occlude considerable practice variations and complexity within organizations. These observations would seem to vindicate critics who argue that the attribution of practice variations to the existence of multiple logics is only an exercise in relabeling institutional phenomena that adds few novel insights to the larger institutional research program and merely reinforces the tendencies toward degeneration (Aksom and Tymchenko 2020; Aksom et al. 2020). Yet other propagators of the institutional logics perspective have insisted that this perspective offers improved explanations of how organizations respond to institutional complexity without over-emphasizing the extent to which such responses are determined by extant institutions (Greenwood et al. 2011; Micelotta et al. 2017; Ocasio and Gai 2020). The institutional logics perspective also goes beyond the view of institutional change as prompted by relatively disembedded agents and conceives of such change as conditioned by the diverse logics in which actors are embedded (Lounsbury 2008; Ocasio et al. 2017). This may be seen as a progressive extension of the institutional research program that avoids the tendencies in much earlier research to attribute institutional change to either the strategic pursuit of vested interests or the relatively rare occurrence of institutional contradictions that cause actors to disembed themselves from extant institutions. However, even propagators of the institutional logics perspective have repeatedly raised concerns that researchers may be tempted to mobilize this perspective in ways that downplay the notion of embedded agency and, thereby, fail to improve on extant explanations of (de-)institutionalization (Thornton et al. 2012; Lounsbury et al. 2021). According to Lounsbury et al. (2021), many researchers either tend to reify extant logics at the expense of the complex processes through which actors reproduce or transform logics or, alternatively, revert to relatively actor-centric accounts of such processes that foreground interest-based conceptions of agency. Similar to much earlier institutional research that brackets the role of either agency or extant institutions in institutional processes, this may reinforce the degenerative tendencies that have long plagued the institutional research program. To summarize the discussion so far, it is clear that the hard core of the institutional research program has remained relatively stable over time even though the program has evolved into a heterogeneous body of scholarship. From the outset, institutional theorists have conceived of organizations as socially constructed entities, whose survival is determined by their ability to achieve and maintain social legitimacy, and have recognized that human agency is a socially conditioned phenomenon and not imbued with unbounded rationality. Even where institutional theorists have sought to introduce a stronger emphasis on agency and change, as in the case of research on institutional entrepreneurship and strategic agency, they have not challenged this hard core although their bracketing of extant institutions might lead to overly actor-centric accounts of (de-)institutionalization. The propensity for such challenges has also been reduced by the reluctance among institutional theorists to subject the assumptions that underpin this core to critical tests and the concerted efforts to advance notions of embedded agency. These efforts have spawned a number of variants of institutional theory that all offer progressive explanations of how the process of (de-)institutionalization unfolds. However, as is clear from the outline of the institutional work and institutional logics perspectives, research emerging within these variants has not always managed to avoid bracketing the role of agency and extant institutions and improve on extant explanations of (de-)institutionalization. Unless a balanced emphasis on agency and institutions is upheld, there is a risk of the institutional research program degenerating as it may lead researchers to revert to either the determinism associated with research on institutional isomorphism or overly voluntarist, actor-centric accounts of (de-)institutionalization. 3. Review of institutional research on MA Scope of analysis and review procedures Extending the discussion of institutional theory, the following literature review examines whether institutional research on MA can be said to constitute a degenerative or progressive research program. To this end, I map the evolution of this body of research from its emergence in the 1980s to the present day.5 To keep the sample of studies under review to a manageable, yet representative, size while controlling for the quality of publication outlets (cf. Hiebl 2021), the search for relevant studies was confined to 11 major accounting journals that have regularly published MA research based on institutional theory.6 The literature search was based on a set of key terms that reflect the evolution of institutional theory described above.7 Based on this search, 163 studies that use institutional theory as their sole analytical lens or alongside other theories were deemed relevant for inclusion in the review. Consistent with the previous outline of institutional theory and the Lakatosian notion that research programs are made up of a set of related and often coexisting theories, the studies were categorized into five clusters (Figure 1).8 An inspection of Figure 1 confirms the Lakatosian notion that the theories that make up a research program do not replace each other through discontinuous breaks, but rather continue to coexist over extended periods of time (Lakatos 1970). Also, even though the number of studies that make up a research program does not tell us much about whether such programs are degenerating or progressing,9 it is notable that all but one of the five research clusters are still growing. The clusters informed by notions of institutional isomorphism and decoupling as well as embedded agency and institutional contradictions have both grown steadily over the past two decades. More recently, research on institutional work and, especially, institutional logics and complexity has grown rapidly without replacing these older bodies of research. The only research cluster that has ostensibly stopped growing is that underpinned by notions of institutional entrepreneurship and strategic agency. Having categorized the studies into the five clusters, each study was read carefully with an eye to the contributions that institutional theory has made to our understanding of MA as an organizational phenomenon and whether accounting scholars also make a more general contribution back to institutional theory (see supporting information in the online Appendix).10 Each study was read independently by the author and a research assistant, who was first trained in the general evolution of institutional theory, after which the readings were compared and any discrepancies in interpretations were resolved through discussions until a consensus solution was reached.11 Although these analytical procedures initially focused on the claimed contributions in each study, they were combined with a more critical reading of whether these contributions should be seen as feeding into broader degenerative or progressive tendencies. Following the incremental conception of what constitutes a theoretical contribution that underpins Lakatos’s (1970) model of scientific progress, the claimed contributions were seen as theoretically progressive if they improved on prior conceptual explanations of how MA is (de-)institutionalized and/or improved on extant conceptualizations of (de-)institutionalization in the larger institutional research program at the time when the contributions were being advanced. The contributions were seen as empirically progressive if they entail empirical support for such conceptual extensions. In applying these criteria, it is important to recognize that each individual study that makes up a larger research program does not have to be both theoretically and empirically progressive for the program as a whole to progress. Some studies may only be empirically progressive in the sense that they provide empirical support for theoretically progressive explanations that have been advanced in prior research. However, since research programs need to be both theoretically and empirically progressive to avoid tendencies toward degeneration (Lakatos 1970), theoretical progressiveness is a necessary precondition for the research program as a whole to progress. Institutional isomorphism and decoupling Similar to the general evolution of institutional theory, the earliest applications of this theory in the MA literature emerged as a distinct counterpoint to functionalist approaches, such as contingency theory, that mainly saw the use of accounting as a means of value-neutral representation and instrumental control. In doing so, accounting scholars began to document how accounting practices underpinned by strong tendencies toward institutional isomorphism are decoupled from operating-level decision-making and control (Covaleski and Dirsmith 1983, 1988a; Ansari and Euske 1987). However, these early advances did not only open up new vistas for accounting scholars by applying institutional theory to particular accounting topics, but also extended institutional arguments. Some of them drew on in-depth, historically informed field studies (Ansari and Euske 1987; Covaleski and Dirsmith 1988a) that went considerably beyond the large-scale, quantitative inquiries that dominated the early development of institutional theory (Tolbert and Zucker 1996) in explaining how novel accounting practices are institutionalized. As such, they recognized that tendencies toward decoupling are far from an automatic response to institutional isomorphism, but an outcome of complex processes that are infused with intricate political dynamics and power struggles. This insight presaged the ambition to imbue institutional analyses with a stronger sense of agency, which has characterized subsequent developments in the larger institutional research program, and can thus be seen as a theoretically as well as an empirically progressive extension of the program at an early stage of its development. These progressive extensions of the larger institutional research program were further developed based on insights that were quite specific to the broader, sociologically informed accounting literature emerging in the 1980s. Some of the early applications of institutional theory in the MA literature examined the roles that accounting practices, such as costing (Ansari and Euske 1987) and budgeting (Covaleski and Dirsmith 1988a), play in communicating broader, societal expectations and institutionalizing such expectations even though accounting practices end up being decoupled from the operating level of organizations. Similar to other strands of sociologically informed accounting research emerging at the time (Burchell et al. 1980; Miller and O’Leary 1987; Hines 1988), this drew attention to the constitutive role of accounting as a calculative practice that is actively involved in the shaping of social realities that, up to this point, had attracted little attention in the institutional research program. Tracing the constitutive effects of accounting over time, Covaleski and Dirsmith (1988a) also documented how the tendencies toward decoupling decreased as new budgetary practices became more firmly embedded and started to reshape operating-level realities. This reinforced the portrayal of decoupling as a more dynamic phenomenon than that found in early variants of institutional theory (cf. Meyer and Rowan 1977; Covaleski and Dirsmith 1988b). The propensity of accounting scholars not only to garner empirical support for institutional arguments in relation to MA, but also offer theoretically progressive extensions to the larger institutional research program, is much less pronounced in subsequent research exploring the influence of institutional isomorphism. The reason for this is partly methodological. The majority of the studies exploring the influence of institutional isomorphism and decoupling are single-site field studies of relatively short duration (Hoque and Alam 1999; Nyland and Pettersen 2004; Pettersen and Solstad 2007), comparative field studies across multiple organizations (Edwards et al. 2000; Lapsley and Pallot 2000; Hussain and Hoque 2002; Kurunmäki et al. 2003; Fallan et al. 2010; Tucker and Parker 2015), archival studies (Eldenburg and Krishnan 2008; Balakrishnan et al. 2010; Eldenburg et al. 2015; Holzhacker et al. 2015; Dai et al. 2018; Firk et al. 2019), surveys (Geiger and Ittner 1996; Cavalluzzo and Ittner 2004; Lapsley and Wright 2004; Johansson and Siverbo 2009a) and mixed methods studies combining survey and interview data (Hoque and Hopper 1994, 1997; Pettersen 1995, 1999; Alam 1997; Malmi 1999; Bhimani et al. 2016). These studies lack a pronounced process focus that is necessary for imbuing institutional analyses with a stronger sense of agency as well as bringing out the constitutive effects of accounting. As such, they offer a view of institutional isomorphism as a largely given, external phenomenon to which organizations respond while bracketing the complex social dynamics through which institutions are reproduced and transformed. Hence, even though this body of research has continued to support institutional arguments and, thereby, contributed to the empirical progressiveness of institutional research on MA as well as the larger institutional research program, it has not offered new insights that make the latter theoretically progressive. Indeed, the majority of the studies exploring the influence of institutional isomorphism only harbor modest, if any, ambitions to advance the larger institutional research program but rather concentrate on contributing to the MA literature. A relatively large and growing number of studies combine institutional and functionalist approaches, such as contingency theory or economic theories, to explore whether these theories complement or compete with each other in explaining the adoption of accounting practices (Hoque and Hopper 1994, 1997; Granlund and Lukka 1998; Malmi 1999; Hussain and Hoque 2002; Siverbo and Johansson 2006; Eldenburg and Krishnan 2008; Balakrishnan et al. 2010; Bol and Moers 2010; Eldenburg et al. 2015; Holzhacker et al. 2015; Firk et al. 2019) and variations in the extent of decoupling (Alam 1997; Geiger and Ittner 1996; Lapsley and Pallot 2000; Cavalluzzo and Ittner 2004; Johansson and Siverbo 2009a). A key argument in several of these studies is that accounting practices are not as homogeneous as some institutional theorists suggest and that practice variations emerge as the constraining influence of institutional isomorphism gives way to functional concerns with organizational performance in at least some parts of institutional fields. At first sight, this insight can perhaps be seen as a theoretically and empirically progressive extension to the accounting literature that refines our understanding of the conditions under which MA is more or less susceptible to institutional influences. However, since functionalist and institutional theories are juxtaposed as analytically distinct perspectives, it does not offer any additional conceptual content to institutional theory per se. From the perspective of the larger institutional research program, it is, therefore, a degenerative, rather than progressive, development. Such degenerative tendencies may threaten the hard core of the institutional research program by leading researchers to smuggle functionalist assumptions back into institutional analyses. This is especially the case in research that combines institutional theory with economic theories (Eldenburg and Krishnan 2008; Balakrishnan et al. 2010; Eldenburg et al. 2015; Holzhacker et al. 2015). Working on the assumption that institutional environments are distinct from technical environments, these studies have conceived of at least some organizations within institutional fields as largely unrestricted by institutional constraints and, therefore, able to pursue functionalist notions of rationality aimed at enhancing organizational performance rather than social legitimacy. For instance, in explaining the varying presence of such rationality, Balakrishnan et al. (2010, 770) argued that “[r]elative to for-profit organizations that typically operate at the technical end, government organizations usually operate in stronger institutional environments” and that this leads the former to prioritize economic efficiency to a greater extent than the latter. Similarly, in their study of the field of health care in the United States, Eldenburg et al. (2015, 169) posited that the economic incentives in place in for-profit hospitals will lead them to “maximize hospital profitability by all means possible” while the institutions that govern non-profit hospitals will “likely mitigate this negative effect.” However, these studies notably fail to recognize contemporary advances in institutional theory, such as the institutional logics perspective, that suggest that all forms of agency and rationality that evolve in institutional fields are conditioned by extant institutions (Lounsbury 2008; Thornton et al. 2012). If they had done so, they might not have degenerated into essentially functionalist accounts of organizational behavior. To some extent, overt threats to the hard core of the institutional research program have been mitigated by the reluctance of MA scholars to reject institutional theory in favor of functionalist approaches or other, competing theories. Even where researchers report empirical findings that partly or wholly contradict the predictions of Meyer and Rowan (1977) and DiMaggio and Powell (1983), they have often sought to nuance such predictions by reconciling their findings with other strands of institutional thought or cognate social theories (Hoque and Alam 1999; Edwards et al. 2000; Kurunmäki et al. 2003; Kasperskaya 2008; Bol and Moers 2010; Fallan et al. 2010; Mättö and Sippola 2016; Chiwamit et al. 2017; Dobija et al. 2019). For instance, Bol and Moers (2010) found little evidence of the forms of decoupling described by Meyer and Rowan (1977), but stopped short of rejecting institutional theory in favor of functionalist arguments and rather explained the institutionalization of accounting practices from a social learning perspective. Such nuancing of institutional arguments can perhaps be seen as a theoretically progressive extension of institutional research on MA as well as the larger institutional research program that helps to refine our understanding of how organizational practices are (de-)institutionalized. However, with a few exceptions (Bol and Moers 2010; Fallan et al. 2010; Chiwamit et al. 2017), it has not been accompanied by any explicit attempts to articulate the contributions to the larger institutional research program and, even where this is the case, researchers do not expound what it is that is specific about accounting as an object of analysis and how this helps to extend this program. There are some exceptions to this tendency to defocalize the specific roles that accounting plays in organizations as a basis for making theoretically progressive extensions to the larger institutional research program. For instance, in contrast to Meyer and Rowan’s (1977) view of decoupling as a matter of separating institutional and functional requirements, Johnsen (1999) advanced the argument that performance measurement practices that display loose couplings between organizational objectives and performance indicators stand a greater chance of success in terms of the use of such indicators than more tightly coupled implementation modes. Similarly, Brignall and Modell (2000) departed from Meyer and Rowan’s (1977) view of decoupling by starting their analysis from an issue that is more specific to the MA literature. Critiquing emerging performance measurement systems, such as the balanced scorecard, for over-emphasizing the benefits of integrating performance indicators into chains of causal relations, they argued that this ignores how the decoupling of diverse performance indicators can help organizations balance competing constituency interests. This approach to decoupling is similar to, but predates, the view of institutional fields as made up of competing institutional logics that are sometimes internalized in organizations but at least partly separated from each other to avoid disruptive conflicts (Thornton et al. 2012; Pache and Santos 2013). However, in comparison with research following Meyer and Rowan’s (1977) view of decoupling, the approaches advanced by Johnsen (1999) and Brignall and Modell (2000) have received much less empirical attention and support in the MA literature (Modell 2003; Rautiainen 2010; van Hengel et al. 2014). Hence, even though their views of decoupling may be seen as theoretically progressive extensions of the institutional research program, they have not generated a sustained stream of empirically progressive work. Also, in contrast to early applications of institutional theory in the MA literature, there is little attention paid to the constitutive role of accounting and how this contributes to the shaping of social realities in research following the works of Johnsen (1999) and Brignall and Modell (2000). Only two studies exploring the influence of institutional isomorphism have rekindled this interest in the constitutive role of accounting (Dambrin et al. 2007; Malmmose and Kure 2021). Based on an in-depth, process-oriented analysis, Dambrin et al. (2007) showed how accounting practices adopted in response to institutional pressures contributed to institutionalize novel, business-oriented ideals in a pharmaceutical company and how the discourses associated with such ideals reinforced their objectification, although they remained decoupled from the subjective identities of individual employees. However, similar to Covaleski and Dirsmith (1988a), Dambrin et al. (2007) cautioned that such states of decoupling may change over time since the constitutive power of accounting may ultimately contribute to the reshaping of subjective identities and align such identities with emerging ideals. This argument was borne out empirically by Malmmose and Kure (2021), who showed that certain accounting practices, which might be expected to be decoupled from operating-level decision-making and control, can create new social realities that may be difficult to change once they are institutionalized and start to transform people’s professional identities. In their longitudinal field study in the Danish health care sector, they observed how the institutionalization of financial control practices reinforced an efficiency-centered rationale that hampered the subsequent implementation of patient-focused quality control practices. Paradoxically, this led the latter control practices to be decoupled from operations despite serving the interests of operating-level employees. Further research of this kind can constitute an antidote to any tendencies to smuggle functionalist assumptions into institutional research. By recognizing that concerns with organizational performance, such as emerging notions of efficiency and quality, are an outcome of the processes through which accounting shapes social realities, researchers are more likely to remain faithful to the social constructivist origins of institutional theory and contribute to a progressive development in institutional research on MA as well as the larger institutional research program. Institutional entrepreneurship and strategic agency Following the early attempts by institutional theorists to imbue research with a stronger sense of agency and change, several accounting scholars started to pay more focused attention to the intricate political processes through which MA practices are (de-)institutionalized (Covaleski et al. 1993; Abernethy and Chua 1996; Collier 2001; Modell 2001, 2004). Several studies follow Oliver’s (1991) framework relatively closely (Abernethy and Chua 1996; Modell 2001; van Helden and Tillema 2005; Grafton et al. 2011; Moll and Hoque 2011; Wijethilake et al. 2017), while others apply this framework more loosely while arguing that organizational responses to institutional pressures do not follow structurally determined patterns but are a result of deliberate, strategic choices on the part of managers or other, dominant organizational actors (Modell and Lee 2001; Modell 2002; Chang 2006; Järvinen 2006; Krishnan and Yetman 2011; Hsu and Qu 2012; Cardinaels and Soderstrom 2013). Other studies mobilize notions of institutional entrepreneurship (Sharma et al. 2010; Gooneratne and Hoque 2016) and de-institutionalization (Seal 2003; Henttu-Ahu and Järvinen 2013; Becker 2014) to examine organizational change processes. However, only a small number of studies have extended this research focus to the institutional field level to examine how MA practices, such as novel forms of performance measurement, are institutionalized at this level of analysis (Østergren 2006; Chang 2009, 2015; Sutheewasinnon et al. 2016). Similar to most MA research exploring the effects of institutional isomorphism, institutional pressures are thus seen as a largely given, exogenous factor and little attention is paid to the constitutive role of accounting in the shaping of social realities over time. Although the studies reviewed above open up a discussion of the role of interests and power in the (de-)institutionalization of MA practices, they emerged and continued to grow during a period when institutional theorists increasingly abandoned the original notions of institutional entrepreneurship and strategic agency in favor of approaches that place the idea of embedded agency center stage. As such, they can hardly be said to represent the state of the art of institutional theory as it has evolved over the past two decades. Even though there is generally some recognition that agents are not completely unconstrained by extant institutions, there is little engagement with institutional theorists advancing more embedded notions of institutional entrepreneurship (Hardy and Maguire 2008; Battilana et al. 2009). Change agents are rather seen as actors with vested interests who are guided by relatively instrumental forms of rationality. Several studies go to some length in describing the intentions of dominant actors and how such actors translate their intentions into various change initiatives, while saying little about how these intentions have been shaped by the institutions in which actors are embedded (Abernethy and Chua 1996; Collier 2001; Modell 2001; Østergren 2006; Ma and Tayles 2009; Sharma et al. 2010; Moll and Hoque 2011; Becker 2014; Palermo 2014; Gooneratne and Hoque 2016; Sutheewasinnon et al. 2016). Hence, much of the institutional influence on human agency is effectively bracketed and, in contrast to research on embedded agency, mainly seen as a force that constrains rather than enables such agency. This lack of improvement on other, contemporary strands of institutional theory reinforces the tendencies toward degeneration, although it has not caused researchers to lean toward functionalist accounts of agency and rationality that threaten the hard core of the institutional research program. Given these degenerative tendencies, it is difficult to see MA research examining notions of institutional entrepreneurship and strategic agency making any major contributions to the larger institutional research program. Indeed, few of the studies in this category make any explicit claims to such contributions (Abernethy and Chua 1996; Collier 2001; Modell 2001; Moll and Hoque 2011; Gooneratne and Hoque 2016) and, even where this is the case, these claims are not very substantial or convincing in light of how this research program has evolved over time. For instance, Modell (2001) claimed to extend Oliver’s (1991) framework by recognizing how efficiency- and legitimacy-seeking concerns are more closely intertwined than she realizes. This can perhaps be seen as a theoretically progressive extension that is reminiscent of the institutionally defined notions of functionality found in the literature on institutional logics (Lounsbury 2008). However, Modell (2001) did not recognize that a deeper appreciation of how institutional logics shape notions of functionality and organizational change processes requires researchers to pay careful attention to how notions of institutional embeddedness evolve over time. Other accounting scholars claim to contribute to the larger institutional research program by directing greater attention to intra-organizational change processes (Moll and Hoque 2011; Becker 2014), the role of individuals (Gooneratne and Hoque 2016) and power and resistance in such processes (Collier 2001), and how organizations might go beyond institutional constraints (Abernethy and Chua 1996). But without being accompanied by a more pronounced sense of institutional embeddedness, it is difficult to see how such advances may be construed as a theoretically progressive extension of this research program. All they are doing is adding empirical observations to a stream of research that is now seen as dated, if not outright degenerative, by institutional theorists. The fact that research exploring notions of institutional entrepreneurship and strategic agency now seems to be in decline may be an indication that MA scholars are somewhat belatedly beginning to recognize and reverse this development. Embedded agency and institutional contradictions In sharp contrast to MA research invoking notions of institutional entrepreneurship and strategic agency, various streams of research that pay much greater attention to the role of embedded agency have emerged since the late 1990s. One of the most prominent streams of research with such a focus is that following the works of Scapens (1994) and, especially, Burns and Scapens (2000). Originating from a slightly different vantage point, which drew extensively on institutional economics, this stream of research soon merged with advances in neo-institutional sociology such as Seo and Creed (2002) that stressed the role of institutional contradictions as a source of institutional change (Burns and Baldvinsdottir 2005; Lukka 2007; Modell et al. 2007). Subsequent studies have also tended to draw on both Burns and Scapens (2000) and neo-institutional sociology and are thus firmly rooted in the latter research program (Siti-Nabiha and Scapens 2005; Hyvönen and Järvinen 2006; Nor-Aziah and Scapens 2007; ter Bogt 2008; Yazdifar et al. 2008; Johansson and Siverbo 2009b; van der Steen 2009, 2011; Boitier and Rivière 2013; Contrafatto and Burns 2013; Mutiganda 2013; Quinn 2014; McLaren et al. 2016; Taylor and Scapens 2016; Ozdil and Hoque 2017; Alawattage and Alsaid 2018). A key insight from these studies is that formal accounting rules, which often emerge from forceful managerial change initiatives, tend to encounter inertia rooted in extant organizational routines and that the human agency involved in change processes therefore needs to be seen as embedded in such routines. Such inertia goes beyond resistance, deliberately exercised by agents with vested interests, and cannot be reduced to strategic agency emerging in response to institutional pressures. Also, in contrast to earlier variants of institutional theory focusing on institutional isomorphism and decoupling, empirical research following Burns and Scapens (2000) does not start from the assumption that organizational inertia will invariably manifest itself in the decoupling of institutionalized accounting rules from the operating level of organizations. Instead, most of this research draws attention to the complex, intra-organizational processes through which change initiatives unfold through detailed, longitudinal field studies. Even where instances of decoupling are documented, this is seen as an outcome of an intricate and recursive interplay between accounting rules and routines, rather than as a deliberate, strategic response to exogenous institutional pressures (Siti-Nabiha and Scapens 2005; Lukka 2007; Nor-Aziah and Scapens 2007; van der Steen 2011). Research following Burns and Scapens (2000) constitutes a theoretically and empirically progressive extension of the MA literature informed by institutional theory that is in line with the growing emphasis on embedded agency in the larger institutional research program from the 1990s (Holm 1995; Greenwood and Hinings 1996; Hirsch and Lounsbury 1997). Following its origins in institutional economics, which emphasizes the inherent embeddedness of all forms of economic agency, this stream of research has avoided the analytical separation between the institutional and technical environments of organizations that characterized early advances in institutional theory and the concomitant tendencies to juxtapose functionalist and institutional theories. By emphasizing institutional embeddedness, it has also eschewed any tendencies to smuggle functionalist assumptions back into institutional analyses. Instead, similar to the institutional logics perspective, it stresses the need to conceive of any notions of functionality or rationality as conditioned by the specific, institutional contexts where accounting practices evolve. A criticism of Burns and Scapens (2000) is that they ignored the wider field-level dynamics that are of central concern to institutional theorists, and how such dynamics interact with intra-organizational change processes (Dillard et al. 2004). This criticism has been at least partly rectified in a growing number of studies exploring how intra-organizational change dynamics are influenced by change initiatives originating at the institutional field level (Hopper and Major 2007; Modell et al. 2007; Boitier and Rivière 2013; Mutiganda 2013; Moore 2013; Alawattage and Alsaid 2018). However, similar to MA research exploring the influence of institutional isomorphism, these studies still view the wider institutional environments that surround organizations as given and thus beyond the influence of individual organizations. Little attention has been paid to the reciprocal relationships between individual organizations and the institutional fields in which they are embedded and how such relationships unfold over time. Turning to the contributions to the larger institutional research program, most of the studies following Burns and Scapens (2000) do not include any explicit claims to such contributions. Yet, this stream of research does draw attention to the constitutive role of accounting in a manner that is not dissimilar to what was documented in early applications of institutional theory in the MA literature (Ansari and Euske 1987; Covaleski and Dirsmith 1988a). Examining how embedded accounting routines are continuously reproduced and transformed, and how such routines are implicated in a recursive interplay with extant and emerging rules that govern MA practices, explains how such practices are not only shaped by but also shape the institutions that evolve in organizations. However, due to its evolution into a relatively self-contained body of research that has not entailed much dialogue with other bodies of accounting scholarship emphasizing the constitutive role of accounting (cf. Justesen and Mouritsen 2011; Miller and Power 2013), it has rarely made this role explicit. Only a small number of studies (van der Steen 2009, 2011; Quinn 2014) have started to engage with such wider bodies of accounting scholarship. Hence, in contrast to early applications of institutional theory in the MA literature, that emerged as a more integral part of a larger corpus of sociologically informed accounting research, there is perhaps a missed opportunity to contribute to a theoretically as well as empirically progressive development in the larger institutional research program. Also, due to the predominant focus on intraorganizational change processes, the processes through which organization-specific accounting practices feed back into and influence the (de-)institutionalization of accounting practices at the institutional field level have rarely been examined in research following Burns and Scapens (2000). This limits the insights into the constitutive role of accounting in the institutional processes affecting larger populations of organizations. Other accounting scholars emphasizing the institutional embeddedness of human agency have tried to make their contributions to the larger institutional research program more explicit by extending research to the role that accounting plays in the intertwining of efficiency and legitimacy concerns (Covaleski et al. 2003; Hopper and Major 2007; Boland et al. 2008; Yazdifar et al. 2008), the institutionalization of novel conceptions of organizational strategy (Modell 2012), the socio-political framing of collective action in organizations (Englund et al. 2013) and institutional fields (Modell and Yang 2018; Becker et al. 2020), and the endogenization of externally initiated public policy reforms in specific organizational contexts (Covaleski et al. 2013; Ahrens and Ferry 2018). These extensions are theoretically and empirically progressive since they all shed light on the role of accounting in relation to themes that, at the time that the research was published, were all under-researched in institutional theory. However, in contrast to research emphasizing the constitutive role of accounting, these contributions are based on insights that are perhaps less unique to the field of accounting. Another area where accounting scholars emphasizing the institutional embeddedness of human agency have contributed to a progressive development in the larger institutional research program is with respect to the conceptualization of power. Several researchers mobilize a view of power as a dynamic and relational, yet institutionally embedded, phenomenon that goes beyond instrumental, interest-driven conceptions of power as a property that belongs to certain actors (Burns 2000; Modell 2005; Tsamenyi et al. 2006), and sometimes extend the discussion of how such research contributes to the larger institutional research program (Modell 2006; Yang and Modell 2013; Major et al. 2018). For instance, based on an empirical study in the Portuguese health care sector, Major et al. (2018) advanced a conceptual framework that explains not only how embedded agents respond to institutional contradictions, but also how the power dynamics that follow from their change initiatives get implicated in a reciprocal interplay with accounting that gradually (re-)shapes social realities. Such research highlights the constitutive role of accounting while feeding into an emerging stream of research that partly overlaps with research on institutional work, and refines the theoretical conception of power in institutional theory (see Lawrence and Buchanan 2017). As such, it contributes to a progressive development in the larger institutional research program that is beginning to steer clear of earlier conceptions of power as exercised by individual and relatively disembedded change agents. MA research exploring the possibilities of embedded agency culminated in a special issue of Management Accounting Research dedicated to the paradox of embedded agency in 2018 (Ahrens and Ferry 2018; Englund and Gerdin 2018; Hiebl 2018; Horton and Wanderley 2018). Although the paradox of embedded agency was explored empirically in several earlier studies within this category of research (Burns and Baldvinsdottir 2005; Englund et al. 2013; Yang and Modell 2013), the contributions to this special issue encompass several conceptual advances with the potential to take MA research as well as the larger institutional research program forward. For instance, Englund and Gerdin (2018) synthesized extant MA research into a framework for explaining how embedded agency aimed at changing extant institutions is enabled, which went considerably beyond the emphasis on institutional contradictions in much prior research on this topic. Similarly, Hiebl (2018) mapped how MA can be used as a political resource to enable embedded agency and argued that more research is required to enhance our understanding of such agency beyond the emergence of institutional contradictions and suggested a number of future avenues for such research. Finally, Horton and Wanderley (2018) combined insights from institutional theory and social identity theory to explain how management accountants can experience identity conflicts and how such conflicts may trigger change in embedded professional roles. These advances all constitute theoretically progressive extensions to the MA literature in that they draw attention to under-researched topics and the possibilities of refining extant explanations of the (de-)institutionalization of accounting. Their conceptualizations of this phenomenon are also sufficiently generic to contribute to a progressive development in the larger institutional research program. However, similar to much prior MA research exploring the role of embedded agency, they entail few explicit claims to such contributions. Unless such claims are more clearly articulated, there is a risk of yet another missed opportunity for accounting scholars to contribute to a progressive development in the larger institutional research program emerging. Institutional work While only representing a small body of research, MA studies using the concept of institutional work offer additional insights into what may cause degenerative and progressive tendencies to emerge. Far from all of this research follows the conceptual underpinnings of this perspective to the letter and it often fails to account for the influence of extant institutions on human agency. Indeed, most MA research using the notion of institutional work concentrates on unpacking the human agency involved in the institutionalization of novel accounting practices without paying systematic attention to how extant institutions condition such agency and the evolution of accounting practices (Goretzki et al. 2013; Hayne and Free 2014; Bhimani et al. 2018; Lagström and Ek Österberg 2020; Aliabadi et al. 2021). The main concern in many of these studies is to classify the individual and collective agency involved in such processes into different categories of institutional work while examining the political dynamics that accompanied institutionalization. However, with the exception of Goretzki et al. (2013), who somewhat peculiarly mobilized the institutional work perspective as an alternative to notions of embedded agency, the authors do not provide any justifications for why the influence of extant institutions was bracketed. Taken together, this reinforces the criticism that the institutional work perspective has been more concerned with unpacking the agency involved in institutional processes than explaining the process of (de-)institutionalization with systematic references to embedded agency (Khagan and Lounsbury 2011; Ocasio et al. 2017; Modell 2022). Even though this does not necessarily mean that researchers revert to functionalist accounts of agency, it constitutes a degenerative development that does not improve conceptually on prior, actor-centric conceptions of (de-)institutionalization. However, some studies using the concept of institutional work have paid more careful attention to notions of embedded agency in explaining how different accounting practices are (de-)institutionalized (Chiwamit et al. 2014; Richardson and Kilfoyle 2016; Gibassier 2017) and have thus avoided degenerative tendencies such as those described above. For instance, Chiwamit et al. (2014) examined the work involved in the institutionalization of MA innovations, such as Economic Value Added, in the fields of Chinese and Thai state-owned enterprises and how such work was conditioned by extant states of field cohesiveness. Such cohesiveness was “defined in terms of how consistent and tightly coordinated the interests clustered around a particular innovation are” (Chiwamit et al. 2014, 148) and was found to have a major impact on how this innovation was embedded in the two fields. Other studies have conceptualized institutional embeddedness in terms of how social class structures and regulatory frameworks, which dominate institutional fields, condition the propensity of various actors to challenge accounting practices (Gibassier 2017) or maintain such practices over extended periods of time (Richardson and Kilfoyle 2016). Taken together, these studies address the concern that much research on institutional work has downplayed or ignored notions of institutional embeddedness and offer novel, empirical insights into how such embeddedness is manifest in diverse institutional fields. Some studies can also be said to have contributed to a progressive development within the larger institutional research program. Chiwamit et al. (2014) extended the institutional work perspective by introducing the notion of field cohesiveness as a novel way of conceptualizing institutional embeddedness and documented how variations in such cohesiveness conditioned the evolution of institutional work across different fields. Similarly, Richardson and Kilfoyle (2016) introduced the notion of routines-as-truce and examined how the institutional work that was implicated in maintaining such truces can be seen as an alternative to institutional isomorphism and decoupling as a source of stability in institutional fields. Both Chiwamit et al. (2014) and Richardson and Kilfoyle (2016) thus offer theoretically as well as empirically progressive extensions to institutional theory. However, in contrast to early applications of institutional theory in the MA literature, most research on institutional work mainly concentrates on the intricate political dynamics and power struggles that unfold around specific accounting practices and has paid little attention to the more specific, constitutive role of accounting in the process of (de-) institutionalization. The only exception to this pattern is Lagström and Ek Österberg’s (2020) study of the introduction of novel forms of accounting calculations to make the notion of social investment operational in the context of social work. Focusing on how such accounting practices became implicated in processes or economization, or the formation of new, calculable realities (Miller and Power 2013), they examined how the institutional work conducted by multiple actors reinforced the creation of such realities. However, as noted above, Lagström and Ek Österberg’s (2020) study falls into the category of research that pays little, if any, attention to the embeddedness of institutional work and they largely bracket questions of how extant institutions influenced the process through which accounting was implicated in creating new realities. Addressing this issue is important to preserve the distinctly institutional identity of research exploring the constitutive role of accounting (cf. Modell et al. 2017) and, thereby, contributing to a progressive development in institutional research on MA as well as in the larger institutional research program. Institutional logics and complexity Following Lounsbury’s (2008) introduction of the institutional logics perspective to an accounting audience, a substantial amount of empirical research has examined how the prevalence of multiple logics gives rise to variations in MA practices across organizations in the same institutional field (Hyvönen et al. 2009; Ezzamel et al. 2012; Kantola and Järvinen 2012; Rautiainen and Järvenpää 2012; Amans et al. 2015; Järvinen 2016; Rautiainen et al. 2017; Rana and Hoque 2020; Damayanthi et al. 2020). The institutional logics perspective has also been used to explain the emergence of diverse, intra-organizational accounting practices that are at least partially separated from each other (Carlsson-Wall et al. 2016; Conrath-Hargreaves and Wüstemann 2019; Kaufman and Covaleski 2019; Lada et al. 2020; Knardal 2020; Lepori and Montauti 2020; Rozenfeld and Scapens 2021) or combined into hybrid practices (Pettersen and Solstad 2014; Busco et al. 2017; Dai et al. 2017; Convery and Kaufman 2021; Kallio et al. 2021). In contrast to MA research ascribing the prevalence of practice variations to the co-existence of functional and institutional pressures in institutional fields, this research sees such variations as firmly rooted in the prevalence of diverse institutional logics. Hence, at first sight, there is little ground for arguing that the institutional logics perspective has failed to contribute to a theoretically and empirically progressive development in institutional research on MA (cf. Aksom and Tymchenko 2020; Aksom et al. 2020). However, in contrast to MA research exploring the notions of embedded agency and institutional work, the processes through which novel accounting practices take shape have received relatively cursory attention. With the exception of a few historically informed field studies (Ezzamel et al. 2012; Conrath- Hargreaves and Wüstemann 2019; Kaufman and Covaleski 2019; Lepori and Montauti 2020; Convery and Kaufman 2021), research has been dominated by comparative or single-site field studies of relatively short duration that yield limited insights into how institutional logics interact with accounting practices over time. This lack of a pronounced process focus has reinforced the degenerative tendencies that propagators of the institutional logics perspective have long cautioned against. Many MA scholars bracket the dynamics through which institutional logics are reproduced and transformed without providing any justifications for why such bracketing took place (Kantola and Järvinen 2012; Pettersen and Solstad 2014; Herremans and Nazari 2016; Järvinen 2016; Busco et al. 2017; Rautiainen et al. 2017; Dai et al. 2017; Knardal 2020; Rana and Hoque 2020; Diab 2021) and, thereby, run the risk of reifying extant logics (Thornton et al. 2012; Lounsbury et al. 2021). Others revert to a view of the agency involved in responding to competing institutional logics as a deliberate, or strategic, phenomenon and focus on how diverse logics are enacted by different actors without paying much attention to how such enactment is conditioned by extant institutions such as the professional backgrounds and identities of individual actors (Hyvönen et al. 2009; Rautiainen and Järvenpää 2012; Carlsson-Wall et al. 2016; Gerdin 2020; Lada et al. 2020; Damayanthi et al. 2020). This leads to a voluntarist view of actors as capable of drawing on diverse logics at will to achieve different organizational objectives. Some studies even combine the institutional logics perspective with Oliver’s (1991) framework for explaining how organizations respond strategically to institutional pressures (Hyvönen et al. 2009; Rautiainen and Järvenpää 2012; Damayanthi et al. 2020) without recognizing the criticisms levied at her conception of agency by propagators of the institutional logics perspective (see Marquis and Lounsbury 2007; Lounsbury 2008). Even though these propagators do not reject the possibilities of deliberate agency, exercised by reflexive human beings, they insist that the capacity for reflexivity is always conditioned by preexisting institutions (Thornton et al. 2012; Ocasio et al. 2017). Analyses of this intricate interplay between reflexive agency and extant institutions have not loomed large in MA research mobilizing the institutional logics perspective.12 Although the tendencies to incorporate notions of strategic agency in research on institutional logics do not necessarily imply that researchers are reverting to functionalist accounts of agency and rationality, there is an imminent risk of this occurring where researchers engage in relatively eclectic theory development by blending the institutional logics perspective with functionalist research approaches. An example of this is Gerdin’s (2020) study of how different institutional logics influenced the use of multiple forms of management control in higher education. Gerdin (2020) adopted the institutional logics perspective as an extension of functionalist research, originally inspired by contingency theory, on how management control systems operate as a set of complementary practices. Although this may, at first sight, seem like a theoretically progressive extension of an established body of MA research, the author severely underplays the institutional embeddedness of the agency involved in the empirical analysis. Even though Gerdin (2020, 2, footnote 3) recognizes that the institutional logics perspective “does not imply that key actors are fully rational and autonomous” and that human agency is always “historically and contextually bound,” the author brackets such notions of embedded agency in the empirical analysis. Instead, Gerdin (2020, 11, emphasis in original) confined the analysis of agency to the observation that key organizational actors may “strategically choose to introduce institutional complexity” in order to “better address particular [control] problems” and “turn institutional complexity into a source of strategic advantage” while paying little attention to the process through which such agency emerged and unfolded. Hence, he is not only introducing a strong notion of strategic agency, but also bracketing the process through which agents reproduced and transformed the logics under examination (cf. Lounsbury et al. 2021). Also, by foregrounding the achievement of presumably performance-enhancing, strategic advantages, rather than legitimacy gains, as the ultimate objective of organizational actors, he produces an account that is more consonant with functionalist, contingency theory-inspired views of organizational behavior than institutional theory. This example shows how problematic even a seemingly deliberate bracketing of notions of embedded agency can be. Even though Gerdin (2020) may not have intended to smuggle functionalist assumptions back into institutional theory, he effectively ends up doing so and his analysis of the agency involved in the design of management control practices does not improve on the explanations of (de-)institutionalization that have long been available in research on institutional entrepreneurship and strategic agency. Although there is thus growing evidence of MA researchers using the institutional logics perspective in ways that reproduce and perhaps even reinforce degenerative tendencies within the larger institutional research programs, others have been more astute in preserving a strong sense of institutional embeddedness when examining how such logics influence accounting practices. Several studies include careful analyses of the agency involved in the (de-)institutionalization of accounting practices without reverting to simplistic notions of interest-driven, strategic agency (Cruz et al. 2009; Ezzamel et al. 2012; Amans et al. 2015; Le Theule and Lupu 2016; Conrath- Hargreaves and Wüstemann 2019; Kaufman and Covaleski 2019; ter Bogt and Scapens 2019; Laguecir et al. 2020; Lepori and Montauti 2020; Rozenfeld and Scapens 2021). Some of these studies have started to connect the institutional logics perspective to the stream of research emanating from the work of Burns and Scapens (2000), while extending the latter body of research with greater attention to how field-level dynamics interact with intra-organizational change processes (ter Bogt and Scapens 2019; Rozenfeld and Scapens 2021). Ter Bogt and Scapens (2019) advanced a conceptual framework explicating how institutional forces at the field level may be translated into intra-organizational accounting practices and how the latter get imbued with a strong sense of context-specific situated rationality that can serve as a bulwark against decontextualized and instrumental conceptions of functionality and rationality. Similarly, Rozenfeld and Scapens (2021) advanced a multilevel analysis that linked the institutional logics evolving at the field level to the individual identities of organizational actors in an attempt to explain how human agency is conditioned by a wider range of institutional forces. Even though these advances do not aspire to making a more general contribution to the larger institutional research program, they offer theoretically and empirically progressive extensions to institutional research on MA by drawing attention to how notions of embedded agency are influenced by the multilevel dynamics that unfold in institutional fields. Other accounting scholars drawing on the institutional logics perspective have advanced theoretically and empirically progressive extensions to the larger institutional research program by continuing to refine our understanding of the constitutive role of accounting (Ezzamel et al. 2012; Le Theule and Lupu 2016; Modell 2019; Kallio et al. 2021). Similar to early applications of institutional theory in the MA literature (Ansari and Euske 1987; Covaleski and Dirsmith 1988a), several of these studies explicitly theorize how accounting, as a calculative practice, is implicated in communicating societal expectations and how this changes the institutions in which accounting practices are embedded. For instance, Ezzamel et al. (2012) did not only examine how the emergent business logic in the field of education shaped the budgeting practices evolving in this field, but also showed that the enactment of budgetary reforms by schools generated both anticipated and unintended consequences that, to a degree, transformed their institutionalized identities. Similarly, Le Theule and Lupu (2016), illustrate how accounting practices, which were embedded in the emergent market logic in the field of publishing, began to encroach on the editorial process in a publishing firm and how this reinforced the marginalization of the previously dominant editorial logic and changed the institutionalized nature of publishing. In contrast to other recent advances exploring the constitutive role of accounting, such as those emerging in research on institutional work (Lagström and Ek Österberg 2020), these studies show that detailed attention to this role can be combined with a strong sense of institutional embeddedness and that such analyses do not need to degenerate into overly actor-centric accounts. However, in analytical terms, they are limited to how institutional logics are transposed from the institutional field level to individual organizations and do not capture the reciprocal relationships unfolding between these levels of analysis. Hence, similar to much prior institutional research on MA, they conceive of the wider institutional environments that surround organizations as given and do not explain how organization-specific accounting practices influence the (de-)institutionalization of field-level practices over time. To address these limitations, Modell (2019) recently advanced a multilevel framework that explains how organization-specific accounting practices, that emerge in response to institutional complexity can fuel diverse framing processes that, in turn, shape the conceptions of organizational performance that are institutionalized at the field level, By conceptualizing organizational performance as a socially constructed phenomenon, Modell (2019) offers a theoretically progressive extension to the larger institutional research program that may provide a powerful antidote to any tendencies to smuggle functionalist assumptions into institutional research. As MA scholars explore how accounting practices are implicated in the shaping of diverse conceptions of performance, they can delve into an important topic that has been under-researched by institutional theorists (David and Bitektine 2009; Lockett et al. 2015). By doing so, they are also likely to resist the temptation to juxtapose functionalist and institutional theories based on the assumption that organizations are either dominated by rational concerns with performance or wider concerns with social legitimacy. Similar to recent research on the influence of institutional isomorphism and decoupling (Malmmose and Kure 2021), Modell (2019) rather views the construction of any notions of organizational performance as intricately intertwined with processes of organizational legitimization. Empirical research extending this line of inquiry can provide a distinct counterpoint to functionalist approaches, such as contingency theory, that have historically dominated MA research exploring how contextual factors affect organizational performance. Institutional Research on Management Accounting 2581 CAR Vol. 39 No. 4 4. Conclusion Adopting a Lakatosian perspective, this paper has asked whether institutional research on MA can be described as a degenerative or progressive research program. While tapping into familiar themes in institutional theory, such as the relative emphasis on human agency and extant institutions in institutional processes. I have reframed the discussion of these themes along Lakatosian lines and examined how MA research has not only been influenced by but also contributed to this larger research program. Admittedly, the Lakatosian model of scientific progress sets a relatively restrictive standard for what counts as scientific progress by insisting that theoretical contributions need to be of an incremental nature and offer improvements on prior explanations of particular phenomena to avoid tendencies toward degeneration. Adopting such a perspective is justifiable in light of emerging concerns about the alleged tendency of institutional theorists merely to repackage extant knowledge, rather than advancing substantially novel insights into organizational behavior (Alvesson and Spicer 2019; Alvesson et al. 2019; Reed and Burrell 2019; Aksom and Tymchenko 2020; Aksom et al. 2020), which have only begun to gain recognition in the accounting literature (Modell 2015, 2022; Modell et al. 2017). However, by taking such a perspective, I have by no means intended to belittle the significant contributions that institutional theory has made to the MA literature and my analysis nuances the criticisms that have recently been levied at its use in accounting research and organization studies. Emerging from a broader corpus of sociologically informed accounting research, institutional research on MA has evolved into a largely progressive research program. This body of research has made important contributions to the MA literature by first establishing institutional theory as a distinct alternative to functionalist approaches, such as contingency theory, that previously dominated research on how MA practices are adapted to organizational environments. It then grew rapidly into a substantial and self-contained research program that has extended institutional arguments to a broad range of accounting topics from the mid-1990s. An especially important stream of research is that emerging from the work of Burns and Scapens (2000). This work spawned a large body of in-depth empirical inquiries into the embedded agency involved in the reproduction and transformation of accounting practices in organizations that is line with the general development of institutional theory over the past two decades. It also reinstated a focus on the complex intra-organizational change processes that featured in some of the earliest applications of institutional theory in the MA literature but that was largely bracketed in subsequent research. More recent advances, drawing on the institutional work (Chiwamit et al. 2014; Richardson and Kilfoyle 2016; Gibassier 2017) and the institutional logics perspective (ter Bogt and Scapens 2019; Rozenfeld and Scapens 2021) have introduced a stronger focus on institutional field-level dynamics that was less salient in research following Burns and Scapens (2000). However, more work is still required into the reciprocal, multilevel dynamics that emerge across different levels of analysis to enhance our understanding of how accounting practices evolving within individual organizations are influenced not only by extant institutions at the field level, but also how such practices influence field-level dynamics. Such multilevel analyses can possibly enhance the opportunities for MA scholars to offer progressive extensions to the larger institutional research program in organization studies, which has been more focused on institutional fields than institutional processes within individual organizations (Greenwood et al. 2014). Although MA scholars have paid less explicit attention to advancing this larger program, there is some evidence of them doing so. Apart from providing extensive empirical support for institutional arguments and thus contributing to its empirical progressiveness, accounting scholars have offered theoretically progressive extensions of this program by inter alia suggesting novel conceptualizations of decoupling (Johnsen 1999; Brignall and Modell 2000) and institutional embeddedness (Chiwamit et al. 2014; Richardson and Kilfoyle 2016) and by advancing important insights into how accounting is implicated in the (re-)shaping of power relations within organizations and institutional fields (Modell 2006; Yang and Modell 2013; Major et al. 2018). The relatively limited efforts of MA scholars to articulate their contributions to the larger institutional research program are perhaps not surprising and resonate with prior assessments of other bodies of accounting research (Lukka and Vinnari 2014). However, it is nevertheless a potential cause for concern in that it may lead to a view of MA research as a mainly applied subfield of organizational sociology that detracts from the progressive extensions that accounting scholars have offered to this program based on insights that are somewhat unique to their field of research. Such research has mainly taken the form of studies emphasizing the constitutive role of accounting in the process of (de-)institutionalization. Concerns with this role featured in some of the very earliest applications of institutional theory in the MA literature (Ansari and Euske 1987; Covaleski and Dismith 1988a) and have been rekindled in a number of more recent studies (Dambrin et al. 2007; van der Steen 2009, 2011; Ezzamel et al. 2012; Quinn 2014; Le Theule and Lupu 2016; Modell 2019; Kallio et al. 2021; Malmmose and Kure 2021). However, in pursuing this line of inquiry, accounting scholars need to take considerable care such that analyses of the constitutive role of accounting are still imbued with a strong sense of institutional embeddedness and do not degenerate into overly actor-centric accounts that possibly over-emphasize the constitutive power of accounting (cf. Modell et al. 2017). As we have seen evidence of in research on institutional work (Lagström and Ek Österberg 2020), there is a real risk of such accounts emerging where researchers bracket the influence of extant institutions. As I have argued throughout the paper, the tendency of researchers to bracket the role of either human agency or extant institutions in the process of (de-)institutionalization and, thereby, underplay embedded agency is indeed one of the main reasons for the persistent emergence of degenerative tendencies in institutional research on MA. This tendency is most notable in research mobilizing the notions of institutional entrepreneurship and strategic agency, which has continued to conceive of human agency as a largely disembedded phenomenon despite the widespread critique of such conceptions in institutional theory. Somewhat ironically, however, it has also continued to plague parts of the research drawing on the institutional work and institutional logics perspectives that were initially devised to address the lack of attention to embedded agency in institutional analyses. Although these tendencies are not unique to the MA literature, but can also be found in the larger institutional research program, MA scholars have also produced degenerative tendencies that have been less extensively documented in this larger program. These tendencies have emerged from the relatively widespread and growing practice of combining functionalist and institutional theories and have effectively led some researchers to smuggle functionalist assumptions back into institutional analyses in a way that threatens the hard core of the institutional research program. To some extent, explicit threats to the hard core of the institutional research program have so far been alleviated by the reluctance of MA scholars to abandon institutional theory in favor of functionalist research approaches. Similar to the general development of institutional theory in organization studies (cf. Davis 2010, 2015), there have been few, if any, outright attempts to disconfirm it and replace it with functionalist accounts of how accounting operates in organizations. However, the tendencies to smuggle functionalist assumptions into institutional analyses arguably represent a more surreptitious threat that may slowly erode the hard core of the institutional research program and undermine its distinct identity. By way of implications for future research, the above discussion underlines the centrality of maintaining a strong emphasis on embedded agency and placing relatively balanced emphasis on human agency and extant institutions as forces explaining the process of (de-)institutionalization. As we have seen, such a perspective has served institutional research on MA well and has contributed to establishing it as a distinct research program in the broader, sociologically informed accounting literature. Although future research can continue to advance this perspective in relation to diverse MA topics, I see particular merit in combining it with an explicit emphasis on the constitutive role of accounting. Such research can not only enable accounting scholars to contribute to the larger institutional research program based on insights that are unique to accounting research, but also constitute a powerful bulwark against any tendencies to smuggle functionalist assumptions back into institutional analyses. By recognizing that accounting does not only reflect preexisting social realities, but that it also plays an active role in the shaping of emerging realities, MA scholars are more likely to remain faithful to the social constructivist origins of institutional theory. However, to continue to further a progressive development, such research would need to pay more systematic attention to how diverse institutions condition the constitutive power of accounting across different contexts. Such research can further incremental theory development as to when accounting becomes more or less influential in the (re-)shaping social realities and can perhaps guard against tendencies to overestimate its constitutive power. To advance a research agenda such as that outlined above, MA scholars are well-advised to adopt a comparative approach to theory development that contrasts the relationships between specific institutions and accounting across diverse contexts. However, researchers need to ensure that such comparative approaches do not encroach on their ability to advance a process-oriented and preferably historically informed understanding of the (de-)institutionalization of accounting. Close attention to the processes through which accounting becomes implicated in the (de-)institutionalization of organizational practices is necessary to trace how it shapes social realities over time. However, to imbue research with a profound sense of how the agency involved in such processes is conditioned by extant and perhaps long-standing institutions, researchers also need to take history seriously (Mutch 2018). Exemplary cases of such research are already available in the MA literature (Ezzamel et al. 2012; Quinn 2014). Research on how MA is implicated in the shaping of social realities may also benefit from nurturing a multilevel perspective that pays close attention to the reciprocal interplay between organizational and field-level practices as it unfolds over time. As noted above, such research is still scarce in the MA literature, but may help in advancing our understanding of how the accounting practices that evolve in individual organizations are implicated in the shaping of field-level dynamics and the institutional environments that surround organizations. A limitation of the Lakatosian model of scientific progress informing my analysis is its relatively narrow conception of the usefulness of academic research. In contrast to other discussions of the usefulness of theory in organization studies (Corley and Gioia 2011) and the MA literature (Malmi and Granlund 2009; Lukka and Wouters 2022), my emphasis has been on scientific rather than practical usefulness. Future discussions of the usefulness of institutional theory to accounting research can be fruitfully extended along the latter lines. The relevance of doing so is underscored by the occasional criticisms of institutional theory for being relatively unknown among practitioners and having little impact, especially on managerial audiences (David and Bitektine 2009). However, when considering the usefulness of institutional theory, it is important not to frame the conception of practitioners too narrowly, but to recognize the attempts that have been made to render it relevant to a wider range of audiences (Lawrence and Suddaby 2006; Hampel et al. 2017). Failing to do so might imply a risk of framing institutional arguments in terms of narrow, managerialist concerns with organizational performance and, thereby, reinforcing the temptation to smuggle functionalist assumptions into institutional theory.

The use of digital technology has great potential for medical research. Particularly, the COVID-19 pandemic has played a central role in accelerating the process of development and sophistication of digital technology in medical research [1,2]. Appropriately employing digital technology can significantly increase both efficiency and efficacy in clinical trials and medical education, which are key to the advancement of medical research [3–6]. More specifically, digital technology can contribute to the development of medical research by improving the efficiency of participant recruitment and retention, health data collection, and data analysis in clinical trials. It also increases the accuracy of the analysis by facilitating communication with the participants of an experiment and reducing the time for data collection. Thus, the development of digital technology is closely related to advances in medical research. In a global perspective, the COVID-19 pandemic provides many countries with an opportunity to recognize the importance of employing digital technologies to respond to global challenge. Research collaboration for the development and distribution of vaccines, the prediction and tracking of confirmed cases, and the sharing of information have become globally critical issues. The role of digital technologies in medicine has received steady attention from scholars from diverse domains. In the field of medicine, researchers have mainly studied how digital technology contributes to the treatment or surgery of patients in specific fields. [7–9]. Scholars from engineering fields have focused on data sharing, protection, and management including the interface, integration, and coordination of data [10,11]. Business researchers have predominantly investigated digital health ecosystem and interactions among various stakeholders from an ecosystem perspective [12–14]. Although these existing studies have offered valuable and diverse insights, these studies have been conducted mainly in the context of specific types of organizations (e.g., hospital, biotech company, etc.) or industries (e.g., biotech industry, pharmaceutical industry, etc.), and national or global-level perspectives have been somehow neglected in this stream of research. However, how digital competency contributes to the innovativeness of medical research has received scant attention and is barely supported by empirical evidence. Based on the awareness of this issue, this study examines how national digital competency impacts the innovativeness of medical research in a global context. As medical research is highly diversified, this study will pay special attention to the effects of national digital competency on the research performance of vital areas in medicine—a medical terminology—that comprehensively refers to four areas: (1) surgery; (2) internal medicine; (3) pediatrics, perinatology, and child health; and (4) obstetrics and gynecology. These areas are major disciplines in the field of medicine, encompassing various subfields. For example, geriatrics is included in internal medicine. (Pediatrics, etc., are included in this category not to distinguish a specific age group, but because babies and children require a fundamentally different medical approach than adults). The reason for focusing on those vital areas is because the research performance of those areas is straightly associated with fatalities through the whole lifespan of human. Knowledge in vital areas particularly becomes significant during pandemics, experiencing a new disease for which epidemiologic data have not been accumulated rapidly, and the death rate is determined within a short time. This is because knowledge about the effects of infectious diseases on body parts is directly related to human lives and forms the gist of coping with unprecedented pandemics. For example, at the beginning of the COVID-19 outbreak, many pediatric patients suffered due to a lack of adaptation period to changes in surgical environments and methods [15]. In addition, many people have experienced unexpected pains due to a lack of knowledge about complications or organ damage caused by COVID-19. In addition, as it takes time for the institution to embrace various effects of rapidly developed knowledge, drastic progress in digital technology in terms of both development and utilization could cause particular social interest in the process of taking advantage of such benefits. This is a social phenomenon that is often accompanied by radically innovative knowledge, and it is not uncommon. Therefore, researchers have recognized the importance of understanding how institutional environments influence the progress of medical research [16–19], but enough empirical evidence still has not been found. Hence, we examined how various institutional factors moderate this relationship. We support our arguments using digital competencies, publications, and diverse institutional data of 63 countries. Our conceptual framework is depicted in Figure 1. The remainder of this paper is organized as follows. First, we discuss how national digital competencies impact the degree of innovation research in medicine, and consequently, how this association is moderated by the government and business environments. Second, drawing on data from multiple databases, we provide the results of an empirical test. Third, we address the implications and contributions of this study. 2. Hypotheses Development 2.1. Digital Competency and Innovative Medical Research Recent medical research requires an integrated system consisting of diverse digital technologies. These technologies typically include the Internet of Things [20], big data analysis [21], AI [22], and blockchains [23]. These digital technologies comprise digital health systems in hospitals or clinics [1]. Clinical trials play a crucial role in medical research by providing researchers with the basic knowledge and data to estimate causality when verifying the efficacy and safety of new therapies and devices as well as prevention and diagnosis [4]. However, conducting clinical trials efficiently is often challenging as inherent inefficiencies are embedded in each stage of procedure, including the identification and recruitment of participants, data collection, and analysis of participants, which result in poor clinical trial participation rates. For instance, cancer-related clinical trials secure only 8% of cancer patients [24]. Furthermore, financial burden caused by the physical distance between the patient and hospital as well as complicated scheduling problems lower the participation rates in clinical trials [25]. However, the experiment environments for clinical research have improved quite slowly over the years, thereby keeping them demanding and expensive. However, employing digital technology provides a stepping stone for improving clinical trials both qualitatively and quantitatively. It allows various clinical trials to be virtually implemented [5] and improves the quality of clinical trials in two aspects. First, the adoption of digital technology contributes to the accuracy of clinical trials by making the key steps in clinical trials more reliable. According to Inan et al. (2020), digital clinical trials comprise three key steps: digital recruitment and retention, which is responsible for the participation and management of participants; digital data collection, comprising data mining and processing; and digital analytics, including data analysis and modeling. By utilizing digital technology, it is possible to increase the participation and communication levels of subjects through social media engagement and online consent. By using wearable and mobile sensing technologies, real-time data collection is possible, and various analyses and modeling are possible using AI. In other words, the digital clinical trial removes various obstacles that act as constraints in the existing clinical trials, thereby enabling the qualitative improvement of the clinical trial itself while saving resources. The second aspect is medical education, which is an essential element for nurturing highly qualified medical researchers, armed with rigorous knowledge. The level of medical knowledge has a decisive influence on the ideas and conduction of clinical trials. However, there is a gap between the education provided by medical institutions and the knowledge required to conduct actual clinical trials [26]. In fact, it is difficult to practice all the theoretically learned medical tests in real life. However, this limitation can be overcome by using digital twin technology, as it can reproduce reality in a virtual space. Students can learn medical knowledge and experience diverse medical situations based on repeatable training such as a 3D surgery simulator using haptic technology. These increasing learning opportunities can have a direct impact on the innovativeness of clinical trials. The study of Chen et al. (2022) presented various application methods, such as medical education training, health and behavior tracking, operation playback and reproduction, and medical knowledge popularization, which can be useful when digital twin technology is used in medical education. In summation, digital technology dramatically improves the quality of clinical trials and medical education that is critical to innovative medical research. Therefore, we propose the following hypothesis: H1. National digital competency will have a positive impact on the innovativeness of medical research. 2.2. Moderating Effect of the Institutional Environment The institutional environment significantly influences the strategic choices of the various actors in both the private and public sectors [27–29]. North (1990) provides a theoretical landscape of such changes, in which formal rules, structured incentives and constraints form the institutional matrix. Although our baseline proposition suggests that national digital competency will, in general, have a positive effect on innovative medical research, we anticipate that these positive effects can be contingent upon the institutional environment. In this context Salman et al. (2014) states that the quality of the public system and management system are the two key elements that should be considered when conducting medical research. Therefore, as investing in and conducting innovative research is inherently uncertain and risky, an appropriate level of stability in regulatory and economic systems is an inevitable environmental factor. In this study, we investigated how government and economic environments moderate the association between national digital competency and the innovativeness of medical research. 2.2.1. Government Environment: Quality of Administration and Legal System In order for new knowledge to be used in a critical sector, such as medical research, the institutions and social systems that enable a country to manage the utilization of such knowledge, a stable market system, infrastructure, and high digital literacy must also be in place [30–33]. According to the OCED—a consortium of advanced countries—for digital technologies to be effectively utilized, the following must be available: infrastructure, public services, and data; effective use of digital data; data-driven and digital innovation; and social institutions such as labor markets and trust in society (https://goingdigital.oecd.org/dimensions, accessed on 11 October 2022). In other words, it is important to have a variety of institutional supports that enable the effective and efficient use of digital knowledge throughout society. Regarding legal and regulation aspects, numerous studies have addressed the quality of the legal system as medical research is closely and sensitively related to personal data, thus the need for ethical protocols, and there remains a responsibility issue for experiments that can critically influence the stability of society [16–19]. Therefore, we predict that countries with a high level of administrative and legal systems will have a more effective utilization of digital technology for medical research. Hence, we propose the following hypotheses: H2a: The quality of the administrative system will positively moderate the association between national digital competency and the innovativeness of medical research. H2b: The quality of the legal system positively moderates the association between national digital competency and the innovativeness of medical research. 2.2.2. Business Environment: Market Freedom and Disparity in Economic Development Market freedom has been considered one of the major institutional environments [34,35]. Market freedom enhances the accessibility to resources [36] and reduces information asymmetry between investors and research teams through improved monitoring systems [37]. It also enables research teams to utilize their resources more [38] and make more explorative projects feasible options. However, if a nation has an unevenly developed economy, a certain area may suffer from low accessibility to medical facilities and information due to the lack of basic digital infrastructure, such as the internet or network, or medical facilities [33,39]. As these areas have limited access to basic medical information, along with low participation rates in clinical trials, they lack the chance to experience improved medical knowledge ([40,41]. Digital technologies such as IoT, AI, and blockchain are applied technologies that can only be operated if basic infrastructure and devices such as computer hardware and wireless networks are available. Further, vulnerable socioeconomic environments lead to low information utilization problems [6,42,43]. Even if the focal region has an adequate level of technological infrastructure that enables people to have high accessibility to information or participate in various clinical trials, limitations in time owing to low income, physical disability, and limited public service due to racial discrimination may reduce the chances of enjoying the benefits of using digital technologies. For example, during the COVID-19 pandemic, low-income residents even in New York City suffered significantly owing to serious health inequalities because of an uneven chance to utilize digital technology [43] Furthermore, given that digital technology is closely related to cutting-edge knowledge, consistent investment is required in its development, diffusion, and market development. In fact, economic level has been cited as a source of various digital disparities, as well as healthcare, and this is clearly observed in the digital health market [42]. A region’s economic level is an important factor influencing the stable development of digital technology as well as market formation. If the growth of the digital health market is difficult, the need for medical research to support it will also decrease. Combining all the discussions presented above, we postulate the following two propositions: H3a: The degree of market freedom will positively moderate the association between national digital competency and the innovativeness of medical research. H3b: The degree of disparity in economic development will negatively moderate the association between national digital competency and the innovativeness of medical research. 3. Methods Using 62 national-level panel data, we investigate the effect of national digital competency on the innovativeness of medical research and how government and business environmental factors moderate that relationship. 3.1. Data and Sample For empirical analysis, we utilize multiple databases. Regarding national digital competency (NDC), we draw the data from theWorld Digital Competency data provided by the International Institute for Management Development (IMD), which is a top-tier global research institute in Switzerland. Since the late 1980s, the IMD’s annual report on national competency based on relevant proxies has been widely acknowledged by researchers in various disciplines [44–46]. To estimate the innovativeness of medical research, we use the data from the Journal and the Country Rank database offered by SCImago, which is an established data-mining and visualization group in Spain that provides a wide range of bibliometric data including journals and citations. The data of SCImago has demonstrated reliability in bibliometric research including top-tier medicine journals, such as Nature and Lancet [47,48]. We obtained the raw numerical values of published medical documents and citation data for each nation and constructed the dependent variable. For control variables, drawing on multiple databases, we collected nation-level data on innovation index health infrastructure, political rights index, globalization index, services sector value-adding, gross domestic product (GDP), government protectionism, science research legislation, and innovation index. We offer the details of these variables in the next section. The final sample of our study comprises 63 countries with 341 nation-year observations between 2015 and 2020. The list of sample countries is shown in Table 1. In total, there are 33 countries from Europe, 8 countries from South America, 2 countries from North America, 14 countries from Asia and the Pacific, 5 countries from Middle East, and 1 country from Africa. Our sample include a wide range of countries, including both advanced economies and catching-up economies. We used 2015 as the starting year because interest in digital health has drastically increased based on the emergence of digital transformation, as illustrated in Figure 1. We use 2020 as the cutoff year as forward citation information generally suffer from the truncation issue [49]. 3.2. Variable Descriptions Dependent variable. To estimate our dependent variable, the innovativeness of medical research, we use the number of forward citations per document published in the fields of Surgery, Pediatrics, Perinatology and Child Health, Obstetrics and Gynecology, and Internal Medicine based on the Journal and Country Rank from the Scimago database. Many researchers address that highly cited research is highly likely to be conducted based on combinations of a broad range of knowledge domains that provide an explorative perspective to researchers and enable them to avoids intellectual lock [50]. Similarly, combinative knowledge from exploratory search can produce more innovative scientific research that ultimately becomes highly cited [51,52]. Therefore, the number of forward citations has been widely acknowledged and employed as a proxy for the innovativeness of research in prior studies [53–55]. We first calculate the total number of published and citable documents in each of the vital areas in medicine and the total number of forward citations that those documents received. Both numbers are aggregated at the nation level. Then, consistent with previous literature, we estimate medical research performance as: Innovativeness of Medical Research i,t,c = Total forward citations i,t,c Citatable documenti,t,c where citable document i,t,c represents the number of citable documents published by country i in medical field c at the time of year t. Total forward citations i,t,c represents the number of forward citations (the document receives after published) of the focal citable document. Independent variable. National digital competency (NDC) is measured based on the digital competency ranking data from the IMD World Competency Yearbook, which offers a comprehensive estimation of the digital and technological level of each nation country by combining statistical and survey data. Moderating Variables. We draw institutional data from the Global Economy database. The quality of the administration system is measured as Government effectiveness index from the Global Economy Database. This measure captures the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the govern ment’s commitment to such policies. Regarding the quality of the legal system, we employ the rule of law index from the same database. This indicator captures perceptions of the extent to which agents have confidence in and abide by the rules of society, the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. To capture the quality of market freedom, we use the business freedom index from the Global Economy database constructing this measure based on theWorld Bank’s Doing Business study. Lastly, the disparity of economic development is measured as the uneven economic development index from the Global Economy database. Control Variables. The research can be affected by the overall innovation environment. Therefore, we used the innovation index of a nation from the Global Economy database. The Global Economy database measured the innovation index (country level) using data from Cornell University, INSEAD, and the World Intellectual Property Organization, which provide an innovation index that comprehensively captures each country’s quality of institutions, human capital and research, infrastructure, and market and business sophistication. We use theWorld Bank’s gross domestic product (GDP) data as our control variable. Economic level has been cited as an indicator of digital technologies in healthcare [42]. We also controlled for policy instruments that might have influenced the quality and application of the research. Based on the IMD National Competitiveness Data, we controlled for the nations’ government protectionism and scientific research legislation (laws relating to scientific research encourage innovation). The IMD also offers the measure of health infrastructure, the degree to which it meets the social needs of the focal society, of each nation. We also control the degree of globalization that may facilitate the innovative research in medicine and political rights index that can potentially influence the credibility of governmental policy. We also control for the portion of the services sector that can affect the business activities in the healthcare industry. Appendix A provides detailed information for variable descriptions regarding measurement and source. 3.3. Models Using unbalanced panel data, we employed a fixed-effect regression model to investigate the effect of national digital competency on the innovativeness of medical research in vital areas and moderating effects delivered by various environmental factors. To control unobserved heterogeneity, we employed a fixed-effects regression model instead of a random-effects model based on the Hausman test [56]. We considered the time lag (two years) between the dependent and independent variables with consideration because the bibliometric information (documents and citations) includes the past two years. (1) IMR1 (S, P, O, I) i,t+2 =   
0i + a1Natioanl digital competency (NDC) i,t +.   
2Controls i,t + ei,t (2) IMR 1 (S, P, O, I) i,t+3 = b0i + 1Natioanl digital competency (NDC) i,t +. 2Natioanl digital competency (NDC) x Institutional environment factors B3Controls i,t + ei,t where   
0i represents country fixed effects and ei,t is the random error. IMR (S), IMR (P), IMR (O), and IMR (I) refer to the innovativeness ofmedical research in Surgery, Pediatrics, Perinatology andChildHealth,Obstetrics andGynecology, and InternalMedicine, respectively. 4. Results Tables 2 and 3 present the descriptive statistics and correlation matrices, respectively. Considering space limitations, we used an abbreviated name of each variable for the correlation matrix. The summary statistics indicated that national digital competency (NDC) was positively correlated with forward citations per document in all vital fields, including Surgery (= 0.30, p < 0.05), Pediatrics, Perinatology, Child Health (= 0.21, p < 0.05), Obstetrics and Gynecology (= 0.20, p < 0.5), and Internal Medicine (= 0.30, p < 0.05). The relatively high correlation among dependent variables could be attributed to their academic relatedness. However, no dependent variable is used in the same regression equation, and hence, multicollinearity was not a major concern in analyses. Table 4 demonstrate the results of the main effect. Hypothesis 1 predicts that NDC will have a positive impact on the innovativeness of medical research in the field of the vital area. In Table 4, there are positive coefficients of Model 1 ( = 3.664, p < 0.001), Model 3 (= 3.826, p < 0.01), Model 5 ( = 3.403, p < 0.05), and Model 7 ( = 5.148, p < 0.01), providing support for Hypothesis 1 with the baseline regression model. These results indicate that a national digital capability positively influences research performance in vital areas. These results are held after employing full model regression in Model 2 (\_ = 2.449, p < 0.05), Model 4 (\_ = 2.606, p < 0.05), Model 6 (\_ = 1.841, p < 0.05),and Model 8 (\_ = 3.251,p < 0.01). Next, we shift our attention to investigate how the main effect is moderated by various institutional variables. Hypothesis 2a posits that the quality of administration will enhance the positive impact of NDC on innovativeness of medical research. In Table 5, the positive coefficients of Model 2 (\_ = 3.052, p < 0.001), Model 4 (\_ = 2.531, p < 0.01), Model 6 (\_ = 3.359, p < 0.01), and Model 8 (\_ = 5.482, p < 0.001) provide support for Hypothesis 2a. Figure 2 provides a plot to understand these results. The plot indicates that the impact of national digital competency (NDC) on the innovativeness of medical research is contingent on the quality of the administration system. Hypothesis 2b predicts that quality of the legal system will strengthen the positive effect of NDC on the innovativeness of medical research. In Table 6, the positive coefficients of Model 2 ( = 2.661, p < 0.001), Model 4 (= 1.668, p < 0.05), Model 6 (= 2.641, p < 0.00), and Model 8 (= 4.720, p < 0.001) offer support for Hypothesis 2b. To aid in understanding these results, we plotted the interaction effects in Figure 3. The slope of the high administration system quality line changes steeply over the high vs. low NDC in the areas of Surgery (IRS) and Obstetrics and Gynecology (IRO), which are more strongly moderated than the others. Hypothesis 3a postulates that market freedom will augment the positive impact of NDC on the innovativeness of medical research. In Table 7, the positive coefficients of Model 2 (\_ = 0.089, p < 0.05), Model 4 (\_ = 0.095, p < 0.05), Model 6 (\_ = 0.101, p < 0.05), and Model 8 (\_ = 0.218, p < 0.01) provide support for Hypothesis 3a. Figure 4 offers a plot to understand these results, and the slope of the high market freedom line increases steeply over the high vs. low NDC in the area of Internal Medicine (IRI) compared with the others. Hypothesis 3b anticipates that disparity in the economic development system will diminish the positive effect of NDC on the innovativeness of medical research. In Table 8, the negative coefficients of Model 2 ( = 􀀀0.702, p < 0.05) and Model 8 ( = 􀀀1.274, p < 0.01) offer partial support for Hypothesis 3b. To help in understanding these results, we plotted the interaction effects in Figure 5. Discussion and Conclusions 5.1. Summary and Implications As seen in the COVID-19 crisis, the development and utilization of digital technologies in medicine is becoming a global issue, not just a matter of individual hospitals or companies. However, while the usage of digital technologies for medical purposes has been studied in various fields including medicine, engineering, and business, the predominant interests of existing studies have been the quality of medical services, data management, and interests of individual institutions in the ecosystem, with micro-level perspectives. The main purpose of our study is to expand interest in digital competency and medical research to the national-level perspective. Accelerated by the COVID-19 pandemic, the development of digital technology has provided significant benefits. Our study used data from 63 nations to demonstrate that digital competency positively impacts the innovativeness of medical research in vital areas: Surgery, Pediatrics, Perinatology, Child Health, Obstetrics and Gynecology, and Internal Medicine, and we discuss how this association varies in different institutional environments. This is an important finding for medical research because producing innovative results is critical to the sustainable progress of the field. In addition, our analysis provides several insights into the use of digital technologies for medical research, which is meaningful to researchers, practitioners, and policymakers. First, while we find a positive association between digital competency and innovativeness in medical research, the benefit of digital competency is notably contingent upon different types of institutional environments. The institutional environment likely health inequality arising from the increased use of digital technology in medicine [3,39]. In other words, although the medical benefits significantly increase with digital technology for individuals (or groups) with the accessibility and ability to utilize such technology, for the remaining individuals (or groups), this change may leave them far behind where they are less likely to leverage these benefits. Hence, future studies can recognize this health inequality and develop digital health technologies to solve social problems. The third important implication is for policymakers and institutions. Our findings support the most vital areas of medicine, but different results were obtained for some regions. For example, Pediatrics, Perinatology, Child Health, Obstetrics and Gynecology are unaffected by economic disparities. This could mean the system weakly affects children’s health or important health issues such as cancer. Therefore, these blind spots should be carefully considered when designing a system. As seen during the COVID-19 pandemic, if a global infectious disease such as a pandemic re-emerges, ultimately, the collective intelligence through research collaboration can make an effective global response the key to cope with the crisis. Countries that adhere to closed systems may be able to keep their distance from some issues that can be potentially problematic, but they will be excluded from many of the great benefits that those collaborations will bring. To make those collaborations effective and efficient, indepth understanding of the digital competency and institutional environment of individual countries must be a precedent. Our study empirically demonstrates that digital competency is conducive to the innovativeness of medical research at the country level. However, sophisticated design for the institutional environment must be concurrently considered to maximize its positive impact. 5.2. Limitations and Suggestions for Future Research Our study has several limitations. Although it measured the innovativeness of medical research using established measures, due to the intrinsic characteristics of medicine, experimentations, services, and practices are critical to realizing the innovativeness of new knowledge. There can be limitations in accurately reflecting the degree of innovativeness using a document-based measure. Future research could conduct an in-depth analysis to investigate how innovative publications are utilized in digital health systems. The scope of the medical research used in this study can be extended. Although this is important because of its direct linkage to mortality, recent medical services and practices require extensive cooperation within medicine, such as anesthesiology or radiology, and across other fields, such as material and biomedical engineering. Hence, further studies should be conducted using more comprehensive data. In addition, while we use the simple slope analysis to depict our moderating effects due to the methodological limitation, future study can employ more sophisticate analysis techniques to investigate the detailed mechanisms of moderating effects. Author Contributions: W.S.: conceptualization,methodology,writing—original draft,writing—review & editing, formal analysis, visualization, B.C.: conceptualization, methodology, writing—original draft, writing—review and editing, formal analysis, visualization. All authors have read and agreed to the published version of the manuscript. Funding: This research was funded by CHA Bundang Medical Center, CHA University. This work was supported by Hankuk University of Foreign Studies research fund. Data Availability Statement: The data that support the findings of this study are available upon reasonable request from the authors. Conflicts of Interest: The authors declare no conflict of interest.

By institutional open-access policies to the results of research (scientific knowledge) we shall understand open-access policies to scientific knowledge endorsed by various institutions (research organizations, universities, funding agencies, publishers, libraries, etc.). At present, these policies have been developed in the United Kingdom for three types of organizations: (1) organizations conducting scientific research, their departments, and associations; (2) research funders; (3) publishers of scientific journals. For all these three types of organizations, registers of respective OA policies have been created. Adoption of these policies will not be possible without launching broader initiatives, statements, and declarations on open access to knowledge in the sciences and humanities and without supporting this process at the highest governmental and intergovernmental levels. We have summarized the currently existing open access initiatives, statements, and declarations in the table below. At present, support for the process of open access to knowledge in the sciences and humanities has been given by governmental and parliamentary committees in many countries of the world, and by powerful organizations, such as UNESCO, OECD, European Commission, European Research Council European, European University Association, etc. In view of the fact that the above OA policies are largely related to regulating the process of self archiving the results of scientific research in institutional OA repositories (archives, libraries, storage facilities) we should explain why the term “self archiving” is more frequently used as compared to the term “archiving.” The fact is that OA repositories (open access electronic archives supported by the standard of the “Open Archive” initiative) contain special directories for researchers (personal areas), where they can create, with the help of special instructions from information robots, collections of their research papers, i.e., it is not the administrator of the OA archive who archives the researcher’s works, but the researcher him/herself. Now we can consider the basic types of institutional OA policies. INSTITUTIONAL SELF-ARCHIVING OA POLICIES These policies, introduced upon the recommendation of the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities are registered in ROARMAP (Registry of Open Access Repository Material Archiving policies, http:// www.eprints.org/openaccess/policysignup). The title of this register says that an institutional self-archiving OA policy (institutional mandate) is registered after creation of an institutional OA repository and its registration in ROAR (the Registry of Open Access Repositories) maintained by the University of Southampton. As of June 1, 2008, the ROARMAP register included 22 mandates of funding agencies, 18 mandates of research organizations and universities (institutional mandates), 2 multi-institutional mandates (associations, international organizations), and 4 departmental mandates, as well as a small number of planned mandates of every class. These mandates are distributed among 24 countries, including Russia (CEMI RAS) and Ukraine (in the Ukrainian Law On the Basic Principles of Information Society Development in Ukraine in 2007–2015”). ROARMAP cites a reference to the order by the director of CEMI RAS on the participation of this institute in the international program Open Access to Research Results (http://www.cemi.rssi.ru/riis/news/ mitial-eng.htm), which obliges researchers to place those research papers that are funded by the state budget into the OA repositories of the CEMI RAS within 6 months. The above-mentioned OA policies are registered in ROARMAP in the form of institutional mandates giving open access to the results of scientific research conducted with the support of public foundations. Note that OA policies are based on the standard code of regulations on self archiving developed by Steven Harnad, Professor at the University of Southampton, which he made public at the presentation of the open access supporting mechanisms at the Berlin Conference (October 2003). A Russian translation of this document can be obtained at: http://users.ecs.soton.ac.uk/harnad/Temp/ declaration.ru.html. In view of its importance for the broad academic community, we shall cite the explanatory note to this document presenting the essence and explaining the open-access mechanism. In the world, there are currently 24 000 journals publishing 2.5 million papers per year on average. In contrast to book authors and journalists, authors of research papers do not expect honoraria for their works. They write their papers for a scientific contribution only, therefore in the paper era these authors were always ready to undertake efforts and bear expenditures to send their papers by mail to their colleagues working in their field, regardless of whether they knew them personally or not. In this era, journals could cover their expenditures for publishing and reviewing through subscription payments only. Universities and research organizations pay for subscriptions so that their scholars can read and use the research results of their colleagues from other organizations. But clearly no institution can afford subscribing to the majority of these 24000 journals and the majority of organizations can afford subscription to a negligible part of the journals, the number of which, during recent decades, has been constantly decreasing with the increase in the prices for the journals, even in the Internet epoch. Naturally, in this situation, research libraries, whose budgets are constantly reduced, cannot subscribe even to the necessary minimum of scientific journals. As a result, as it was in the paper era and is still true today in the epoch of the Internet, each of the 2.5 million papers published every year loses the greatest part of its potential readers because it is not accessible to them. This means that each paper loses the largest part of its potential scientific contribution. In the book-publishing era, this loss of scientific contributions was inevitable, but in the Internet epoch it can be avoided. In Steven Harnad’s opinion, there are two ways to prevent this loss. New on-line open-access journals cover their expenditures, not through subscription campaigns but through fees paid by authors or their sponsors (institutions or grants). But at present there are only 1500 such journals, which publish only 5% of the 2.5 million papers issued every year. For the remaining 95% of all papers, which are distributed via subscription by 22500 journals, a solution has been found which could soon put an end to closed access and the loss of scientific impact. In this case, each author is offered the opportunity to self-archive an additional copy of each paper written by him/her on the network server storage of his/her university or research institute (institutional open-access electronic archive), which would support the standards of the Open Archives Initiative, so that this paper could bring its fruit to the benefit of potential users all over the world, whose institutions cannot afford to pay for subscription to the official version of the journal. More than 92% of the journals represented by their publishers have already given their official permission for the self archiving of papers in these online storage sites. The only thing that universities, institutes, and funding agencies should do is to adopt the respective policies on open access to research results and require that these results not only be published, but also that an additional copy of each preprint or postprint paper be deposited in the respective open-access electronic archive. As a result, the progress of science will be enhanced and not senselessly constrained, as it is now. Coming back to the ROARMAP register analysis, note that OA policies are described there in a concise form. Below we shall present the contents of a number of OA policies that have been adopted or are planned for adoption in four countries, which are ambitious from the standpoint of science. On May 23, 2007, the Brazilian Parliament passed a law regarding the dissemination of scientific research results. Its first article reads that all universities and research institutes should mandate the creation of institutional OA repositories, where the results of scientific research are to be stored. In India, the National Knowledge Commission requires that all the research papers published by Indian scholars and funded from public resources be archived in the standard OA format on the authors’ personal websites, and thus be in open access. Later, when a national academic OA portal will be created, these scientific publications will also be placed there. The Chinese Ministry on Science and Technology is preparing a mandate on electronic archiving of scientific research results. The Middle East Technical University (Turkey) obliges all its researchers to place copies of all their papers, whether published or under review, as well as Masters or Ph.D. degree theses, on the university OA repository and takes on the responsibility of encouraging and supporting these authors to publish their papers in OA journals. OA POLICIES OF FUNDING AGENCIES While the ROARMAP register run by the University of Southampton gives a short description of research funders’ OA policies, in the specialized SHERPA JULIET register run by the University of Nottingham these policies are described comprehensively and presented in the form of three subpolicies: 1. Open Access Archiving: this requires open free access to published papers or the reviewed paper version (postprint), although the publishers’ time embargos nullify the online access to these papers. 2. Open Access Publishing: this requires publishing in OA or hybrid OA journals to accelerate the process of disseminating the results of scientific research. 3. Data Archiving Policy: this requires archiving the primary data within a certain time frame. By June 1, 2008, 32 funding agencies from the United States, United Kingdom, France, Germany, Italy, Ireland, Switzerland, Belgium, Canada, and Australia registered their policies in SHERPA JULIET (in ROARMAP there were 22 adopted and 4 planned policies). Surveys conducted in the United Kingdom in 2005 showed that 15% of all authors have been already self archiving their papers in OA repositories, but in case employers and funding agencies require the self archiving of papers, 95% of the researcher are ready to do it and 81% will do it willingly. 1 Moreover, for institutions that have adopted self archiving mandates, the percentage of these authors is approaching 100% (http://eprints.ecs.soton.ac.uk/l 10061). PUBLISHER OA POLICIES ON SELF ARCHIVING Within the framework of the SHERPA POMEO project (at the University of Nottingham), three policies were originally proposed regarding self archiving of papers in OA repositories (http://romeo.eprints.org/ publisher.html): (A) the pale-green policy allows the self archiving of preprints (the author’s version of the paper prior the first contact with the referee and even the publisher); (B) the green policy allows self archiving of postprints (the final author’s version of the paper after refereeing); (C) the gray policy means that self archiving is formally not supported. At the beginning of 2008, these policies, within the framework of the same project, were transformed into broader policies also covering copyright issues, viz., publisher copyright policies and self archiving (<http://vvww.sherpa.ac.uk/romeo.php>). When refining these self-archiving policies, four colors were introduced instead of three, as well as restrictions and terms for postprint self archiving: (1) the green policy means that an author is allowed to self-archive preprints and postprints, with no time restrictions (embargos) for the latter, yet on certain conditions related to copyrights; (2) the blue policy means that an author is allowed to self archive postprint versions with no restrictions but on certain conditions (the self archiving of preprints is not supported); (3) the yellow policy means that the author is allowed to self archive preprints, while self archiving of postprints is not supported or subject to time restrictions (embargo) and copyright terms; (4) the white policy means that self archiving is not formally supported and it is necessary to make requests to the publisher to get permission for self archiving for each article. The descriptions of publisher’s policies on copyrights occupy a fairly large place, and in the ROMEO guide they are cited as hyperlinks to the respective sections of the publisher’s website. The above terms, written for postprint self archiving, form an integral part of the publisher’s copyright policy. All the OA journals allow self archiving of preprint and postprint author’s versions of the papers and, according to the ROMEO classification, fall into the “green” category. It is important to note that in terms of content, postprint is analogous to the published paper; however, in term of appearance it is not similar, since the publisher reserves the rights for the arrangement of typesetting and formatting. In fact, this means that the author cannot use the pdf file created by the publisher and, therefore, should created his/her own pdf version of the paper for placement in an OA archive. At the same time, some publishers insist, in contrast, on authors using the publisher’s pdf file because they want to see their professional pdf file on the Internet to reserve and promote their house style. Let us cite a number of publisher self-archiving policies from the SHERPA Romeo guide. I. Interperiodica: (1) Blue publisher. (2) Self-archiving status for preprints is not defined. (3) The author can self archive postprints on the following conditions: (3.1) Only on personal and institutional sites and on the sites of non-profit organizations. (3.2) Publisher contract and source of the publication should be cited. (3.3) A link to the publisher’s website should be given. II. Elsevier, excluding the journal Cell Press : (1) Green publisher. 2) Authors can self-archive preprints. (3) The author can self archive postprints on the following conditions: (3.1) The source of the publication should be cited. (3.2) A link to the journal’s website should be cited. (3.3) The publisher’s pdf file of the paper cannot be used. (3.4) In some journals, papers can obtain open access status after paying additional expenses. (3.5) Authors working in the system of the National Institute Health (United States) can place their papers in Pub Med Central in 12 months after publication. III. Springer: (1) Green publisher. (2) Author can self archive preprints. (3) The author can self archive postprints on the following conditions: (3.1) On their personal site. (3.2) In the institutional repository and the funding agency’s repository (server) in 12 months after publication. (3.3) The publisher’s pdf file of the paper cannot be used. (3.4) Reference should be given that the original publication of the paper is accessible at www.springerlink.com (3.5) The source of the publication should be cited. (3.6) A link should be given to the journal version of the paper. IV. Blackwell Publishing: (1) Yellow publisher. (2) Authors can self archive preprints. (3) Author can self archive postprints with a time restriction (embargo from 6 months, in rare cases, from 24 months) on the following conditions: (3.1) Only on their personal, institutional, or disciplinary server. (3.2) The server should be noncommercial. (3.3) The publisher’s pdf file of the paper cannot be used. (3.4) The publisher’s copyright and source should be cited with the following phrase: “the definitive version is available at www.blackwell- svnergy.com” (3.5) A link should be given to the journal version of the paper. (3.6) In some journals, papers can obtain openaccess status after paying additional expenses. At present, there are two large registers of institutional repositories, one being supported by the University of Southampton (Registry, ROAR, http://roar. eprints.org/, 1071 OA repositories as of June 1, 2007), and the other by the University of Nottingham (Directory of Open Access Repositories, DOAR, http://www. opendoar.org/, 1148 OA repositories as of June 1, 2007). Russian institutional OA repositories are most fully represented in the first register of OA repositories, and Ukrainian institutional OA repositories are to the same extent presented in both registers (the first register includes five OA repositories and the second one six). Regarding academic OA journals, we should say that as of June 1, 2008 their register (Directory of Open Access Journals, DOAJ, http://www.doaj.org) run by the University of Lunda (Sweden) includes, 3401 OA journals, among which there are 14 Russian journals and 10 Ukrainian ones. At the same time, Brazil has 321 OA journals; India, 97; Turkey, 92; Pakistan, 38; Iran, 33, and the United States has 731 OA journals. Apart from the aforementioned three types of institutional OA policies, in our view, it is necessary to develop OA policies for academic libraries (to search for a compromise between traditional subscriptions and online access to academic periodicals, regulations and priorities concerning digitization of library resources, etc.) All the four types of institutional OA policies are presented in the figure. In view of the fact that in Post-Soviet territories the process of providing open access to scientific research results remains sluggish we suggest organizing efforts on invigorating this process at seven levels: (1) the global-ideological level: to join the international open access initiatives and declarations (the Berlin Declaration and Budapest Initiative); (2) the regional (transnational)-ideological level: to adopt a series of open access initiatives and declarations by analogy with the transborder interuniversity Belgorod Declaration on open access to scientific knowledge and cultural heritage; (3) the national-ideological level: to adopt a series of open access initiatives and declarations, for example, within the framework of national academies of sciences, national university associations, etc.; (4) the national-political level: to adopt a series of government and parliamentary mandates authorizing open access to the results of scientific research performed with support from public funds; (5) the institutional-political level: to adopt institutional OA policies for organizations conducting scientific research (institutional mandates), publishers of academic periodicals, funding agencies and register them in the respective international directories; (6) the national-technological level: to develop programs and large-scale projects on creating a national network of OA repositories and OA journals, as well as to upgrade software and information search systems based on open code; (7) the institutional-technological level: to constantly create institutional OA repositories and OA journals, along with their registration in the respective

Studies on human are imperative for medical progress and have expanded our understanding and capability to treat serious diseases and entities. However, research with humans needs to take into account the ethical dimensions of the reasons for running an experiment and the proper procedural steps to ensure that the results reflect good science. Protecting human participants in research is our top priority and has been given great consideration in the ethical conduct of research because the exact risks and benefits of research are uncertain. “All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood” (Article 1 of the United Nations Universal Declaration of Human Rights). These rights have often been ignored in public perceptions of human research. Beginning in the seventeenth century, the scientific revolution brought about a method of investigation using controlled observation and reporting of result to the public as proof. The numbers of participants involved in early experiments were small and most often included the researchers themselves or their families. The most typical and famous example of this was when Edward Jenner tested a smallpox vaccine on his son and on the neighborhood children in the early modern times. The progression to the current status of protecting human research participants has been the consequence of historical events in the twentieth century. There have been many groundbreaking events that have affected the public's perception of human clinical research. The history of human subject abuses, scandals, tragedies and the responses to them are shown in Fig. 1 in chronological order. 9420 (Singapore), Japanese doctors conducted live experiments with dissection, dismemberment, and bacteria inoculation on prisoners of war. They induced epidemics on a large scale, with an estimated 3,000 to 200,000 Chinese, Korean, Mongolians, and Allied civilians becoming infected [1,2]. Many prisoners were killed, directly or indirectly, by these experiments. After the war, the Supreme Commander of the Allied Powers in Japan, Douglas MacArthur, gave immunity in the name of the United States to Shiro Ishii and all members in exchange for protecting the results from the Soviet Union. No formal investigation or trial took place in association with the Japanese experiments. In the meantime, the Nazis were placing victims in vacuum chambers with low air pressure and a lack of oxygen in order to determine the health effects on pilots at extremely high altitudes. Subjects were immersed for hours in tubs of ice water, fed nothing but salt water for days, and experimented upon with techniques for battlefield medicine. At the end of the war, 23 Nazi doctors and scientists were put on trial in Nuremberg from December 9, 1946 to August 20, 1947 for the unethical treatment of concentration camp inmates, who were often used as research subjects with fatal consequences. Seven were sentenced to death. A set of standards known as the Nuremberg Code was used for evaluating and judging the defendants. The nuremberg code and the declaration of Helsinki The Nuremberg Code comprises such principles as informed consent and absence of coercion; properly articulated scientific experimentation; and beneficence towards experiment participants [2]. The code states that : 1) Voluntary informed consent is essential without any coercion; 2) Human experiments should be designed and based upon prior animal experimentation; 3) Expected scientific outcomes should justify the experiments; 4) The experiment should be conducted only by qualified scientists; 5) The experiment should be conducted in a way that avoids all unnecessary physical and mental suffering and injury; 6) There should be no expectation of death or disabling injury from the experiment. In 1953, the World Medical Association (WMA) was provoked to make drafts that would apply the Nuremberg Code to the practice of human experiment in the medical community. Known as the Declaration of Helsinki, it was an expansion upon the Nuremberg Code and was first adopted in 1964. It has been revised several times (1975, 1983, 1989, 1996, 2000 and most recently in 2008) according to the modern ethical theory and current clinical and research practice. A prominent point of difference from the Nuremberg Code was the flexibility of the conditions of consent, which was 'absolutely essential' under the Nuremberg code. Research was permitted without consent where proxy consent, such as that of a legal guardian, was available. The Declaration of Helsinki introduced the concept of an independent committee, which evolved into the institutional review board (IRB) system used in the US [1]. The Declaration of Helsinki focuses on a systematic approach, including IRB review, unlike the Nuremberg code, which focused on the responsibility of the individual scientist, had no legal enforcement and was applied only to non-therapeutic clinical research. The Declaration of Helsinki is an important document in the history of research ethics as the first significant effort of the medical community to regulate research itself. It forms the basis of most subsequent documents and is now widely accepted as the cornerstone document of human research ethics. The Beecher article Dr. Henry K. Beecher, an anesthesiologist, reported 22 studies describing violations of serious ethical principles in the New England Journal of Medicine in 1966 after the publication of the Declaration of Helsinki [1,3]. This article sparked a debate on research ethics in the US. His examples were not cited simply to blame individuals but with the hope that it would call attention to abuses, in order to correct them. The experiments that Beecher cited demonstrated ethical abuses. Here are two examples: number 7 - this study on cyclopropane anesthesia and cardiac arrhythmia involved 31 patients. Carbon dioxide was injected into the closed respiratory system until cardiac arrhythmias appeared. Toxic levels of carbon dioxide were achieved and maintained for considerable periods, causing various pathologic arrhythmias. Number 17 - live cancer cells were intradermally injected without consent into 22 chronically ill, debilitated non-cancer patients for a study of immunity to cancer (Jewish Chronic Disease Hospital Case, 1963). The physicians "did not wish to stir up any unnecessary anxieties in the patients" who had "phobia and ignorance" about cancer, so they did not tell the subjects that the injection contained cancer cells. The Tuskegee study and the Belmont report The Tuskegee syphilis study was an infamous clinical experiment undertaken by the U.S. Public Health Service, which would later become the Centers for Disease Control and Prevention (CDC), to study the natural progression of untreated syphilis between 1932 and 1972 in Tuskegee, Alabama. The study was designed to demonstrate the need for establishing syphilis treatment programs by investigating the effects of untreated disease. A total of 399 poor, rural black men were enrolled, under the impression that they were receiving free health care from the U.S. government. Select research participants were given free medical care, meals, and free burial insurance. However, they were never told they had syphilis, nor were they ever treated for disease. In spite of the wide use of penicillin as a curative treatment for syphilis by 1951, treatment continued to be withheld from the research subjects. The announcement of the Declaration of Helsinki in 1964 had no effect on the study. Jean Heller, an Associated Press reporter, published a story about the study in the New York Times and the Washington Star on July 25, 1972. The public reaction was great and Senator Edward Kennedy held hearings about these experiments on human subjects. The syphilis study was stopped, and treatment was given to the survivors in 1973. President Clinton officially apologized to the research subjects and their families in 1997. Congress passed a National Act in 1974 creating the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The National Commission published the so-called “Belmont Report” in 1979, which is a landmark of ethical principles in human research. The three fundamental ethical principles for using any human subjects for research are: 1) Respect for persons: protecting the autonomy of all people and treating them with courtesy and respect; this is applied in the informed consent process. Researchers must be truthful and conduct no deception; 2) Beneficence: incorporating the philosophy of "Do no harm" while maximizing benefits for the research project and minimizing risks to the research subjects is applied through risk/benefit assessments; 3) Justice: ensuring reasonable, nonexploitative, and well-considered procedures are administered fairly and equally and applied to the selection of research subjects. These principles are comprehensive and are stated to understand the ethical issue. The three principles cannot always be applied so as to solve beyond dispute particular ethical problems, however, and provide an analytical framework that will guide the resolution of ethical problems arising from research involving human subjects. Today, the Belmont Report continues to be an essential reference for institutional review boards (IRBs) and remains the basis of human subject protection regulations. Human radiation experiments Eileen Welsome revealed to the Albuquerque Tribune in 1993 that researchers injected plutonium into unknown subjects to study the effects of the atomic bomb under government sponsorship [2]. In 1944, President Clinton formed the Advisory Committee on Human Radiation Experiments (ACHRE) to investigate human radiation experiments and decide upon ethical and scientific standards for evaluating these events. The Advisory Committee found that several thousand governments had sponsored human radiation experiments, intentionally releasing radiation on hundreds of occasions from 1944 to 1974. The Nuremberg Code and the Declaration of Helsinki appear to have been disregarded during Cold War radiation experiments. Thalidomide tragedy Thalidomide was approved in Europe as a sedative drug in the late 1950s and sold in a number of countries around the world from 1957 until 1961. It was withdrawn from the market after being found to have caused birth defects in 10,000 to 20,000 children. The FDA had not approved the drug but U.S. physicians had studied its safety and efficacy. The drugs had the side effects of shrinking blood vessels and disrupting the normal development of the vessels, affecting development of the arms and legs. It was extremely damaging to the fetus if taken in the first trimester of pregnancy. In the congressional hearing with Senator Hubert Humphrey from 1959 to 1962, it was found that many people who were taking the unapproved drugs were neither informed that they were being given an experimental substance nor asked for their consent. This led to the passage of the Drug Amendments sometimes referred to as the Kefauver-Harris Amendments. Since the episode with thalidomide, researchers have been required to inform subjects of a drug's experimental nature and to receive their consent before beginning trials. The Milgram study The Milgram experiment (1963) was a series of social psychology experiments conducted by Yale University psychologist Stanley Milgram after reading about the Nazi Holocaust. The study was intended to measure the willingness of participants' obedience to the authorized person who instructed them to perform acts that conflicted with their personal conscience. Volunteers were recruited for a study of "memory and learning". The volunteer was to play role of "teacher" and was required to ask the "learner" questions and administer punishment via an electric shock when the learner gave wrong answer. In reality, there were no electric shocks to the learner, but they pretended to receive an electric shock. Two-thirds of the volunteers were persuaded by the investigator to administer shocks up to the highest level of 450 volts. Upon completion of the experiment, the investigator explained the deception. The focus of Mailgram’s investigation was the psychological stress induced by the experiment upon the volunteers, the deception involved and the lack of true informed consent. As a result of this controversial study, the conditions of deception in human research were limited, and now need careful IRB approval. Hepatitis in retarded children Experiments were designed to track the development of the viral infection of hepatitis and subsequently to test the effects of gamma globulin in preventing or ameliorating the disease from 1963 through 1966 at the Willowbrook State School, a New York State institution for mentally retarded children [4]. The subjects, all children, were purposely infected with the hepatitis virus; early subjects were fed extracts from the stool of infected individuals and later subjects received injections of more purified virus preparations. This Hospital did not admit new patients after 1964, unless their parents consented to the experiment. This case drew public condemnation because of the perception that parents and their children were given little choice about whether or not to participate in research and for performing an experiment on either a normal or a mentally retarded child when no benefit can result for the children. San Antonio contraceptive study and Tearoom trade study San Antonio contraceptive study: In 1971, an oral contraceptive study was conducted on 70 poor Mexican-American women to evaluate the efficacy of different kinds of female contraceptive pills. A number of indigent Hispanic women, who had no way of getting contraceptives, came to a clinic seeking contraceptives. They agreed to participate in a study to determine the side-effects of contraceptives. The randomized half received oral contraceptives and the others a placebo. The two halves were switched in the middle of the study. They were not informed that they were subjects of this kind of research or that they might receive inactive medication. As expected, there were high numbers of unplanned pregnancies in the placebo group; ten of the 76 participants became pregnant during the study. Tearoom trade study: Anonymous male homosexual encounters in public restrooms (a practice that was known as "tearooming" in US gay slang) were studied in a controversial 1970 Ph.D. dissertation and book titled "Tearoom trade: a study of homosexual encounters in public places" by Laud Humphreys. Humphreys, as social scientist, acted as a watcher outside public toilets where people grouped to engage in anonymous homosexual activity. He copied down license plate numbers and other identifying information, which he used to get the names and addresses of over 100 men who had been involved in 50 sex acts (mostly oral sex). He then personally visited their homes to interview them about their milieu and family life. Many subjects were living with a family in a situation where it would be upsetting to disclose their homosexual activity. At no time were the subjects informed that they were participating in a study about male homosexuality. In his published reports, the level of detail was such that the identification of some of his subjects was revealed. Death of Jesse Gelsinger Jesse Gelsinger, an 18-year-old volunteer, was the first person publicly identified as having died in a clinical trial of a gene transfer experiment in 1999 [5]. He suffered from ornithine transcarbamylase deficiency, an X-linked genetic disease of the liver, the symptoms of which include an inability to metabolize ammonia - a byproduct of protein breakdown. He was injected with an adenoviral vector carrying a corrected gene to test the safety of the procedure and died four days later, despite not being sick before the experiment. The principal investigator and the University of Pennsylvania shared in a private startup company that owned the technology used in the experiment. The main issue in this research was conflict of interest (COI). Moreover, investigators did not pay attention to animal data indicating the possibility of adenovirus-induced liver failure and the possible harm to Jesse's already abnormal liver function. Investigators did not use the IRB-approved consent form and had reported instances of mild liver toxicity in previous participants as adverse events. Death of Ellen Roche Ellen Roche, a healthy 24-year-old volunteer in an asthma study, died in 2001 because she inhaled hexamethonium, a medication used for treating high blood pressure in the 1950s and 60s [5]. She developed a cough and her condition worsened over the next week until she was put on a ventilator with progressive multi-organ failure. She was a technician from the Johns Hopkins Asthma and Allergy Center who volunteered to participate in a study designed to provoke a mild asthma attack in order to help doctors discover the reflex that protects the lungs of healthy people against asthma attacks. She died about a month after taking part in the study. Although both a National Institute of Health (NIH) and the IRB had approved the study, hexamethonium was not approved as medication by the Food and Drug Administration (FDA). A federal investigation found serious problem with IRB reviews at the University and accused the IRB of failing to take proper precautions. The IRB did not follow federal regulations and all federally funded research was suspended. Other universities were shocked and began to strengthen their IRB committees. The public expressed outrage at this case, which was readily understandable. The culture of possibly putting coercive pressure on Asthma and Allergy Center employees to participate was pointed out as a grave mistake. International Ethical Guidelines for Human Subjects The Council for International Organizations of Medical Sciences (CIOMS) in Collaboration with the World Health Organization (WHO) guidelines The CIOMS (http://www.cioms.ch/) is an international, non-government, not-for-profit organization established jointly by WHO and UNESCO in 1949 to serve the scientific interests of the general international biomedical community, and has been active in dispersing guidelines for the ethical conduct of research. The international ethics guidelines created in 1993 by CIOMS and updated in 2002 for biomedical research including human subjects were intended to guide investigators from more technically advanced countries when conducting research in developing countries. The guidelines were intended to supplement alleged omissions from the Nuremberg Code and the Declaration of Helsinki, particularly when applied to crosscultural study. The CIOMS guidelines take into account cultural differences in ethical standards. The CIOMS 21 guidelines (15 in the original report) address issues including informed consent, standards for external review, recruitment of participants, and more. The guidelines are general instructions and principles of ethical biomedical research, and have been revised to account for the latest ideas and practices, such as the Declaration of Helsinki. The International Conference on Harmonization- Good Clinical Practice (ICH-GCP) guidelines The ICH (http://www.ich.org/) is composed of expert working groups from the pharmaceutical industry and regulatory authorities in the European Union, Japan and the United States, as well as those of Australia, Canada, the Nordic countries and the World Health Organization (WHO). The goal is to discuss the scientific and technical aspects of drug registration and published guidelines for GCP in response to the increasingly global face of drug development, so that the benefits of international harmonization for better global health can be realized worldwide. The objective of the ICHGCP (Geneva: 1996) guidelines is to provide a unified standard for the European Union (EU), Japan and the United States to facilitate the mutual acceptance of clinical data by the regulatory authorities. Thus, any country that adopts this guideline technically follows this same standard. Clinical studies should be carried out according to International Conference on Harmonization (ICH)/WHO Good Clinical Practice standards. This worldwide GCP document offers standardization for clinical trials of drugs. Standards for the design, conducting, analyzing, monitoring, auditing, recording, and reporting of clinical trials provide assurance that the data and reported results are credible and accurate, and that the rights, integrity, and confidentiality of trial subjects are protected. Ethical and scientific quality standards for designing, conducting, recording and reporting trials that involve the participation of human subjects ensure that the rights, safety and well-being of the trial subjects are protected. GCPs are consistent with the ethical principles originated in the Declaration of Helsinki. The ICH topics are divided into four categories (Q: quality topics, S: safety topics, E: efficacy topics E6 (R1: Revision 1) - Good Clinical practice, M: multidisciplinary topics) and ICH topic codes are assigned according to these categories. The ICH-GCP includes the following sections: (Section 1): Glossary, (Section 2): The Principles of ICH-GCP, (Section 3): Institutional Review Board/Independent Ethics Committee (IRB/IEC), (Section 4): Investigator, (Section 5): Sponsor, (Section 6): Clinical Trial Protocol and Protocol Amendments, (Section 7): Investigator's Brochure, (Section 8): Essential Documents for the Conduct of a Clinical Trial. ICH-GCP, therefore, embraces all aspects of all clinical trials. KGCP (January 1, 2000) was completely revised to harmonize with ICH-GCP regarding standards for clinical trials of drugs in Korea; compliance with KGCP during clinical trials is inspected for all investigations. Task and Responsibilities in Human Subject Research Institution The Institution has the responsibility to comply with the laws and guidelines regarding oversight of all human research activities, especially when the research involves vulnerable people [6]. It also has the responsibility of educating investigators on ethical issues, scientific truthfulness, preventing misconduct and conflicts of interest. The institutions are required to have 1) ethical (IRB) review of protocol and informed consent, 2) administrative review of proposals, contract and grants, 3) scientific peer review [6]. Ethical review: By compliance with the law and guidelines, the institution can guard the rights, safety and welfare of research participants. The IRB must review the following requirements in order to give approval to research: 1) the risks are rational and minimized in relation to the anticipated benefits to the subjects based on a risk/benefit analysis; 2) the choice of subjects is equitable; 3) informed consent is obtained from each potential subject or a legally responsible representative unless waived in harmony with the law and guidelines. This should be documented on the consent form; 4) when subjects are likely to be vulnerable to coercion or undue influence, additional safeguards are needed; 5) appropriate monitoring and observation with continuing review should be scheduled when collecting data to ensure the safety of the subjects, protect the privacy of participants and to maintain the confidentiality of data. The purpose of the IRB is to ensure that the investigator complies with the protocol and to demonstrate that the trial is necessary and that the risk-benefit ratio is acceptable by reviewing key trial documents to ensure that the subjects’ rights and well-being are protected. Administrative review: the research institution generally ensures that proposals and allied budgets are in compliance with the law and institutional policy including IRB review where suitable. If the researcher has a conflict of interest, the institution should make a decision as to whether the conflict can be managed. The research institution has usually established a Conflict of Interest (COI) committee to avoid and/or to minimize potential conflicts under the instruction of institutional policy. Scientific peer review: scientific review should examine the soundness and worth of the hypothesis, the procedure to prove the hypothesis and the appropriateness of the methods to be used. It is unethical to expose subjects to unnecessary risks and sample size justifications must be back up based on the expected results and statistical significance. When the IRB plays the dual role of conducting the scientific review, scientifically qualified experts must be added to the IRB, or the IRB should establish a subcommittee for supporting scientific review. ICH-GCPs provide protection for human volunteers and ensure the accuracy and reliability of data generated in the course of clinical trials. Compliance with these standards is a public pledge that the rights, safety and well-being of clinical trial participants will be protected. GCPs cover obtaining informed consent, documentation, reporting adverse events and proper record keeping. Investigator The welfare and safety of research subjects is ultimately the responsibility of the investigator. The researcher thus shares responsibility with the research institution and sponsors. Investigators must be properly qualified to conduct the research and studies must be suitably designed to produce valid results. Investigators are responsible for ensuring that research is conducted according to the research design as approved by the IRB [4,6]. Good and professional judgment is required throughout the research process to guarantee the protection of study subjects. Investigators must protect and respect the personal dignity and autonomy of the research volunteers by obtaining informed consent before a person agrees to participate in a study. Subjects are protected from harm by study proposals that maximize anticipated benefits and minimize possible risks. The benefits and burdens of research are reasonably distributed. Protecting subjects and achieving scientific progress are not exclusive and not conflicting. The principal investigator can delegate study-specific task and responsibility to other team members including subinvestigators, the Clinical Research Coordinator (CRC), as well as a variety of professionals, statistician, laboratory technicians and administrative staff. Studies should be conducted according to the protocol (study design) that the IRB approved. This is the duty of an investigator in amenability with the regulations. The protocol is a formally written document detailing how the research is to be conducted. The institution policies, guidelines and law state the items that must be included in the protocol and informed consent. The study procedures and inclusion/exclusion criteria are to be evaluated and checked while the protocol is in its draft form. The investigator ought to decide upon the feasibility of recruiting volunteers with/ without advertising prior to approving a study. All changes to the protocol must be agreed upon by the IRB and sponsors before execution. Investigators should document and clarify deviances from the protocol. The detection of major or repeated noncompliance with the protocol can result in closing of the study or even ineligibility as an investigator. Ethical Issues Ethics in clinical design Researchers and IRB members must carefully inspect and bear in mind the details of research design protocol such as randomization, blinding, and the problem of placebos as controls and assessment of risks and benefits. The distinction between research and treatment The ethics of research and therapy are fundamentally different. However, clinical research and therapy both provide medical care and are performed by physicians with similar interventions of treatment in the clinical setting [2]. Experimental interventions and the best proven therapy should appear equally effective. Physicians commonly conduct clinical research and medical therapy as intimately connected. The purpose of clinical medicine is to provide optimal medical care for individual patients; it is ethically governed by the principle of therapeutic beneficence and nonmaleficence. On the other hand, clinical research is not a therapeutic activity devoted to the personal care of patients. It is carried out to answer a scientific question with the aim of producing knowledge that can be generalized and applied to future patients. The clear demarcation between research and therapy becomes blurred when physician-investigators view patients as subjects in practice. Physicians and patients commonly fail to appreciate the distinction between research and therapy because of the similarity in the physician and patient relationship, especially with regard to the setting out of innovative or non-validated therapies. To be sure, the risks need to be assessed by physicians and patients and they must weigh carefully the options of standard treatment and research intervention, of course with the informed consent of the patient. Clinical equipoise and randomized clinical trials (RCTs) RCT is a study design that randomizes whether the participants are given treatment or placebo for the sake of eliminating prejudice. RCTs are ethical only in conditions of "clinical equipoise" being assured. Random selection of participation can yield scientifically convincing data for use in future patients. However, critics of RCTs say that individual is determined not by the participants' physical needs and personal value but by the statistical requirements of the study design. Randomization to get data for future patients sacrifices benefits for the present patients. RCTs violate the physician's duty of giving the most appropriate treatment to their patients. One way of solving this problem is to obtain fully informed consents of the participants. Small losses in some patients might be ethically tolerated as long as the patients are not exposed to unnecessary risk. RCTs are ethically permissible using a standard of clinical equipoise in the context of non-life threatening therapies. Serious problems remain, however, in clinical equipoise that can easily be upset. So long as the study intervention is balanced, RCTs are acceptable. Placebos in clinical research RCTs are well recognized as the most desirable type of study to evaluate a new treatment, but many clinical trials are concerned about the use of placebos as controls. Placebo controls are intended to ascertain the authentic effectiveness of a treatment while eliminating various disturbing factors and to determine the actual therapeutic efficacy of a new treatment. If researchers wish to test a new treatment in the absence of a known effective treatment, the use of a placebo is usually problematic and unethical. Comparisons of new drugs to current standard medications and comparisons to placebos are different. The latter comparison conflicts with the Declaration of Helsinki, which requires that any new method be tested against the best existing prophylactic, diagnostic, and therapeutic method(s). Placebos can have their own powerful ambiguous effects. Comparing against placebos is not the same thing as testing against nothing. A lack of difference between a new drug treatment and the standard treatment does not necessarily mean that the new drug is effective. The new drug and the standard treatment could both be effective or both be ineffective. The standard treatment might be generally effective, but lose its effect in a particular situation. The FDA considers placebo controls to be the gold standard of measuring diagnostic or therapeutic efficacy because they rely on statistical significance in judging the efficacy of the new drug. It is likely that placebo studies will continue to be used. However, they should be used with caution so that patients do not face unnecessary pain or disease on account of a medical experiment in keeping with the ethical use of placebos in any experiment. The ethics of phase I research The main purpose of Phase I trials is to determine the highest tolerated dose of a new drug in humans, with the hope of gathering information that may help patients in the future. Human studies, especially phase I cancer trials, bring about much tension and conflict between the goals of science and those of clinical care, bringing special challenges to IRB review. Almost all Phase I studies are executed on normal human volunteers to determine the level of toxicity and pharmacologic effects of receiving higher doses of a drug on a small number of participants. However, studies that are conducted on sick patients, such as trials of cancer drugs, can be extremely controversial because the drugs are too toxic to be administered to a healthy volunteer. This category of patients is seriously ill and highly vulnerable. These individuals are designated to participate in phase I oncology trials for the good of society with no premeditated benefits and need special protection. Sometimes they are under the misconception that the trials are designed to help them [5]. Consent documents should detail the purpose of this trial and indicate that the dose will be increased until the patient gets extremely sick. Moreover, it is impossible to predict the side effects that the patient will experience because the study is designed to push the dose of the study drug until toxicity is unacceptable. Despite this, most participants think that the main purpose of trial is to make them better. Information including the purpose, risks and benefits of the study should be provided to make clear the distinction between research and patient care. Standardized wording should be required on these consent documents. Participant recruitment Clinical trials should be conducted with the willingness and generosity of those who serve as human participants. Recruitment is almost inevitably time-consuming, expensive, and requiring of the investigator's realistic determination of its feasibility prior to performing the trials. Many patients still have the idea that clinical trials are treatment, especially when they have serious disease. Investigators should guard against exaggerating the benefits of research and should ensure realistic assessments of the benefits and risks before volunteering their patients to become subjects. Concerns prior to participation are the fear of receiving a placebo instead of the active drug, as well as the risky side effects. The fact that research participants are supererogatory volunteers means that investigators and physicians should sustain heavy responsibilities not to violate their trust. People should be selected to make sure that the burdens and potential benefits are equitably dispersed. It is ethically justified to exclude those at greater risk of injury. Therefore, after careful selection of subjects best able to answer the scientific questions and to understand the risks and potential benefits posed by that particular trial, participants are identified, recruited and enrolled according to their eligibility criteria. The scientific and ethical basis of including women and minorities in clinical research are that many have begun to see access to clinical research and to test drugs as an advantage rather than a burden from which people should be protected. Some even saw their participation in the research as not only beneficial, but as essential to their medical care and their chance of survival. On the other hand, once recruitment and enrollment of participants with appropriate inclusion and exclusion criteria have been decided, one controversial problem is the amount of money to pay. Payment should be prohibited, although compensation for expenses may be ethically permitted. Informed consent The voluntary consent of the participant in a clinical trial is now an indispensible part of human research. The process need to include the three key components of information, understanding and voluntary agreement, in order to be ethically suitable. The firmest foundations for the requirement to seek consent are based upon the ethical principle of respect of persons described in the Belmont Report. These imply that individuals should be treated as self-ruling agents and that person with diminished autonomy should be protected. Participating subjects will be treated as an end and not merely as a means to another's end, based on Kantian terms. However, informing the prospective subject that a clinical trial will be at least in part a means is a consent issue in human research that differs from practice. Only emergency and therapeutic concession exceptions are allowed in the context of medical practice. In cases of emergency or life-threatening situations, informed consent can be impossible to get and can sometimes cause postponement of asking the consent of the subject or permission. There is continued controversy over deferred consent as privileges [7]. The therapeutic exception to withholding information is when disclosure would be harmful to the patient's interest or well-being. The subject might be invited to consent to incomplete disclosure with the promise of full disclosure at the termination of the research. Fully informed consent is an ideal goal that we can never achieve, but we must attempt to reach it. Competence and comprehension to reach an enlightened decision is the domain of controversy. Many studies involve unreal or uncertain benefits and the subject's participant represents only a societal good. We need to provide subjects the opportunity to choose what is best for themselves in order to gain their trust while also taking into account the ethical issues of consent. International research A vital issue in international research is exploitation in developing countries. In most developing countries, obtaining voluntary and informed consent is problematic, making it difficult to conduct studies [8]. Many trials that make use of impoverished populations in developing countries violate the most fundamental understanding of ethical attitudes. However, researchers insist that doing research with placebo-controlled studies in developing countries is at least equivalent to the standard of care in these countries, which consists of unverified regimens or no treatment at all. It is now ethically acceptable to most that researchers working in developing country have a responsibility to provide treatment that conforms to the standard of care in the sponsoring country, and, when possible, to resolve the double standard between developing and developed countries. Cultural relativism or community beliefs cannot be used as a justification for violating universal human rights. There must be a core list of human rights that must be protected despite local distinctions in their superficial features. Ethical standards in medicine similarly cannot be relative. The force of local customs or law cannot justify abuses of certain fundamental rights, and the right of self-determination based on informed consent. When researchers from developed countries collaborate on studies performed in developing countries, it is important to stick to these fundamental principles to avoid ethical imperialism and to justify studies. There is an enormous amount of research to be done in developing countries, with their diverse and large populations and the burden of public healthcare that has yet to be solved. A truly international effort is needed to relieve the populations that have suffered so dreadfully. A collaborative effort will be required to conduct ethically and scientifically sound research that yields solid results. Other issues Remaining issues include special populations, genetics research, stored human biological specimens, human embryos and stem cells, drug challenges and drug washout studies, research with communities, scientific misconduct, behavior of clinical investigators, conflicts of interest, research with secondary subjects, tissue studies and records review, and behavioral research issues [4]. These issues are not presented here due to lack of space, but need to be debated. They have not been excluded here because they are any less important than those discussed above. Criticisms to the IRB System and Suggestions IRB review is the main body of research supervision, making IRBs the key protectors of human research participants. However, concerns have been raised about the adequacy of IRB review. In spite of the roles and responsibilities of IRBs, the fact is that many are overloaded, understaffed and faced with a variety of skeptical criticism. Many IRBs are lacking the resources and staff to carry out the hefty task of reviewing research [9,10]. IRBs have acknowledged a number of criticisms for their performance: 1) the monitoring function of IRBs ongoing research is not fulfilled on their role for annual review, consent, adherence to protocol, and data integrity. Auditing and quality assurance programs serve an important preventive role; 2) both free standing commercial review boards (noninstitutional review boards), which are financially dependent on their client, and academic IRBs, the members of which are inclined to accept the studies of their colleagues, have conflicts of interest inherent in their structure. The independence and integrity of both types of IRBs should be secured to avoid problems; 3) multi-center trials by different IRBs cause delays and inconsistencies in IRB review. Exempted or expedited review at another site might be considered to eliminate duplication of effort and to reduce workload when the same study is fully reviewed at some local IRB. The central IRB model with facilitated review process could be a reasonable way to lessen the burden on local IRBs; 4) IRBs pay out too much time reviewing and revising consent forms. Usually consent forms are written at the reading level of a college graduate, and different IRBs in multi-center trials may produce inconsistent consent forms; 5) a review of the scientific benefits of the trial is often beyond the scope of the IRB. Accreditation of IRBs may be an effective approach to improving quality, as an indicator of superiority in human subject protection. The Association for the Accreditation of Human Research Protection Programs (AAHRPP) carries out voluntary accreditation of IRBs requiring self-assessment, site visits, and evaluation. Electronic and structured forms are also suggested to reduce paper work and expedite the review process. Conclusions Biomedical research has made remarkable advances over the past century; as a result, ethics in clinical research is of more concern than ever before. There was little public dispute over the ethics of biomedical research until the 1960s, when scandals appeared to erupt worldwide and were opened to the public (Fig. 1). There have been many responses to these scandals including recognition of the need for standards and guidelines in the ethics of clinical research. The growing necessity for ethics in clinical research has raised concerns related to controversial issues in the processing of the formal mechanism known as the IRB. There exist various perspectives in special topics with or without consensus. This paper first introduces historically evoked scandals and responses, and then identifies key ethical issues and insights, with topics limited by space constraint. Selected debates are intended as a guide to the ethical issues confronted by physicians and researchers. Research ethics is an essential part of good research practice to protect participants in clinical studies. It is our optimistic belief that these challenging issues will be resolved through a consensus in the future. It is also my hope that this review provides an idea of the ethical framework to those investigators and anesthesiologists who will need to meet the challenges of changing patterns of research circumstances.

In recent past, efforts to promote research integrity (RI), which broadly means ensuring the performance of research to the highest standards of professionalism, rigor, and in an ethically robust manner, abound (Adebamowo 2007; Caplan et al. 2018; Gutierrez et al. 2017; Hyder et al. 2007). Such efforts include RI capacity development. Collectively, RI capacity development aims to provide an evidence base for improved practices and policies to maintain or improve research quality. In this paper, we focus broadly on policies, structures, and capabilities in Ghana, a lower middle-income country (LMIC) in West Africa, to learn how research in the Ghanaian setting is produced with integrity. RI, as defined above, encompasses both subject protection and maintaining high scientific quality, including prevention and detection of research misconduct and the maintenance of valid scientific records. Because health research is likely to translate rapidly into clinical practice and directly impact lives, ensuring a high degree of integrity in health research is of particular importance. While significant effort has been invested in building capacity for human subjects protection—through the establishment of Institutional Review Boards (IRBs) in LMICs (Glickman et al. 2009; Hyder and Rattani 2014; Nchinda 2002; Petryna 2009; Strosberg et al. 2013; WHO 1996), very little focus has been placed on ensuring RI (Ana et al. 2013). IRBs, also referred to herein as Research Ethics Committees (RECs) and Ethics Review Committees (ERCs), among others, serve to protect the rights of, or interests of human subjects/study participants. Whether or not IRBs effectively carry out this level of oversight is not routinely evaluated in Ghana. A recent assessment of local IRBs revealed that although their standard operating procedures closely align with international standards/recommendations, unique challenges (such as absence of a formal system to evaluate the activities of the IRBs, high protocol-to-committee member ratio and therefore inability to provide timely feedback on research protocols) remain (Boateng 2019). Other crucial steps central to RI occur largely outside of oversight: how the protocol is implemented, data acquisition, privacy protection, avoidance of misconduct, and the adequacy of data analysis. Other quality control mechanisms such as institutional oversight and peer review, which largely involves examining the finished product, sometimes are unreliable. Over the years, efforts at defining, detecting and developing interventions for preventing research misconduct (RM) and promoting integrity in research have been made (Marusic et al. 2016; Resnik and Master 2013; Anderson and Kleinert 2013; Lee 2012; Resnik and Shamoo 2011; Farthing 2001). RM that violates data integrity is defined in part by the US Federal Policy on the subject as “fabrication, falsification, or plagiarism in proposing, performing or reviewing research, or in reporting research results” (Code of Federal Regulations 2011). In addition to deception in proposing, carrying out, or reporting results of research, others have broadened to definition to include “deliberate, dangerous or negligent deviations from accepted practices in carrying out research. For instance, failure to follow established protocols or adhere to established ethical principles if this failure results in unreasonable risk or harm to humans, other living organisms or the environment and facilitating of misconduct in research by collusion in, or concealment of, such actions by others. It also includes any plan or conspiracy or attempt to do any of the above” (University of Sheffield Human Resources 2020). Fanelli and colleagues’ study of retractions/corrections “supports the notion that scientific misconduct is more likely in countries that lack research integrity policies” (Fanelli et al. 2015). Also important is the absence of a science culture in which mutual criticism is hampered and where publications are rewarded with cash (Fanelli et al. 2015). In a related study of problematic image duplication, academic culture and peer control were found relevant to achieving integrity (Fanelli et al. 2018). National policies appear to be a baseline protection. Policies depending entirely on whistleblower complaints have been found in the US to be associated with significant underreporting of suspected fabrication, falsification, and plagiarism (FFP) (Titus et al. 2008). A combined meta-analysis, largely of US and European studies, found a reported incidence of RM behaviors of about 2% (Fanelli et al. 2018). In a convenience sample of investigators from several Middle Eastern universities, 28.6% of respondents self-reported fabrication and falsification, which was significantly lower among individuals who held a degree from a Western university (Felaefel et al. 2018). A small (N = 158) survey in medical schools and hospitals in India found 57% of respondents knowing of another who had altered or fabricated data in order to get it published (Dhingra and Mishra 2014). Digging deeper into the subject, Kingori and Gerrets (2016) explore the “morals, morale and motivations for data fabrication” in the sub-Saharan setting. They note that fabrication and falsification of data among nonscientists collecting data in the field has been underexamined. Rohwer and colleagues surveyed corresponding authors of Cochrane reviews working in LMICs, finding many describing a lack of clear policies regarding RM and a lack of RI offices in their institutions (Rohwer et al. 2017). Nearly 80% of the survey respondents noted that guest authorship occurred at their institutions, while 40% indicated that colleagues had not in the past declared conflicts of interest (COI) and that researchers were uncertain what COI is and how it may influence their research. Much focus is currently being placed on reproducibility and questionable research practices (QRPs). QRPs generate research findings that may be neither valid nor replicable. Concerned about the low reproducibility of many scientific findings (Ioannidis 2005), many scientists, journals, and funders are looking to RI codes or innovations to improve the reliability and robustness of research. Marcus Munafò describes the practice of “p-hacking (where researchers exploit analytical flexibility to obtain a statistically significant finding), and then present these results as if they were anticipated a priori (also known as HARKing—Hypothesizing After the Results are Known)” (Munafò 2016). Open science, a movement to encourage openness in science through open access publication and open data archiving (Attwood and Munafo 2016), is being promoted as a means to improve the reproducibility of published work. A study of NIH-funded scientists found 70% reporting QRP behaviors (Fanelli et al. 2015). QRPs identified by Sacco and colleagues from surveying individuals with at least one active NIH grant during 2016 include: concealing data, results, methodology, or sources of financial assistance; failing to disclose all relevant COIs; withholding publication to please a sponsor; providing a biased peer review to delay publication and drawing strong inferences from statistically significant but under-powered studies (Sacco et al. 2018). Such QRPs are largely unregulated and leave enforcement to peer review, mentors, and scientific self-correction over time. No comparable statistics exist for Ghana. In Nigeria, Okonta and Rossouw (2013) determined the prevalence of RM in a group of researchers and factors associated with the practice. They view their data as cause for serious concern and call for prompt intervention. Interventions to reduce RM, they authors note, should proceed from measures that contain both elements of prevention and enforcement. Training on research ethics and RI must be integrated into the curriculum of undergraduate and postgraduate students, while provision should be made for in-service training of researchers (Okonta and Rossouw 2013, 2014). Yet another global effort at promoting research honesty is the TRUST Project (European Commission Infocentre 2018). The project is leading far-reaching efforts to improve adherence to high ethical standards in scientific research around the world. The TRUST project’s Global Code of Conduct for Research in Resource-Poor Settings aims to ensure communities, research participants and local resources in scientific studies are treated with fairness, respect, care and honesty. We concur. Whilst codes, tools, and innovations are necessary, the capacity needed to deploy them is essential. Existing efforts of building enduring RI in the sub-region are few (Adebamowo 2007; Hyder et al. 2007). Capacity building requires assessment of current levels of these elements of RI and of professional and/or state regulatory systems to support them. We first describe selected examples of international partnerships; second, the research regulatory environment and the ethics apparatus in Ghana; and, then, a project jointly undertaken by Ghanaian research leaders and scientists in collaboration with New York University School of Medicine’s Division of Medical Ethics to assess the perceived adequacy of current institutional practices, opportunities and incentives in supporting RI. Selected Research Capacity Development Successes and Challenges in Ghana There are some powerful examples of partnerships to create, sustain, and lead research capacity and infrastructure development in pursuit of better health for Africans and Ghanaians. The Global Emergent Pathogens Treatment Consortium (GET) was established by academics, scientists, clinicians, and civil society to facilitate development of infrastructure to support a harmonized African approach to health crises on the continent. GET’s focus is in biobanking and biosecurity in laboratory settings. GET, as a registered entity, has operational offices in several countries, including the West African states of Ghana and Nigeria (Abayomi et al. 2016). The H3Africa consortium, begun in 2010, while funded by NIH and the Wellcome Trust, is African-led. H3Africa aims to develop capacity for genomic studies. By 2017, the consortium had funded 26 research projects in 27 African countries, including Ghana. H3ABioNet will assist with data quality control and analysis (Ramsay et al. 2016). In particular, this consortium has placed a strong emphasis on capacity development in research ethics, particularly community engagement and informed consent (Munung et al. 2017). The West African Network of Excellence for TB, AIDS and Malaria (WANETAM), funded by the European and Developing Countries Clinical Trial Partnership, is engaged in local capacity building to address these diseases. Previously underreported, standardized international drug susceptibility testing is necessary for an adequate surveillance system and treatment. During this project, Ghana’s Korle- Bu Teaching Hospital became accredited as the country’s national TB reference laboratory. This credential moved West Africa toward independent, local research and competitiveness for international TB trials (Gehre et al. 2016). The National Institute of Mental Health (NIMH) has funded Collaborative Hubs for International Research in Mental Health, including in Ghana, to increase the evidence base for mental health interventions in LMICs. Emerging researchers receive seed grants and research mentoring for capacity building (Pilowsky et al. 2016). NIH has funded the Sickle Pan-Africa Research Consortium (SPARCo) on sickle cell disease, using genomics to catalyze discoveries in Africa. Ghana has leadership of a data coordinating center; development of capacity and infrastructure is currently underway. The science is to be led from Africa not only to develop locally effective, evidence-based solutions but also to produce work of global importance (Makani et al. 2017). Other partnership examples include the Ghana-Dutch Research Collaboration (GDRC), and Ghana’s participation in the European Developing Countries Clinical Trials Partnership (Atelu et al. 2016), World Health Organization task forces, and Special Programs for Research and Training in Tropical Diseases (Warsame et al. 2016). In addition, Ghana has participated, in partnership with Novartis, in capacity building for clinical pharmacology research. Locally conducted research in this field is essential to safeguarding drug efficacy, safety, and quality, and aims to understand the influence of genetic diversity on drug response and disease susceptibility (Gutierrez et al. 2017). In their “Mapping of health research institutions in Ghana: Landscaping and comparative analysis,” Seddoh and colleagues offer some lessons from the GDRC (Seddoh et al. 2015). Initiated in 2000, the GDRC was a product of the Health Research Project (HRP), which involved the Netherlands Development Assistance Research Council (RAWOO) and the Ghana Health Service (GHS) Health Research Unit. The HRP aimed to conduct research that would assist the health sector to improve health care in Ghana, with the potential end users of research as the target of any research conducted. The program was financed by the Netherlands Directorate General of Development Cooperation. A Joint Ghanaian-Dutch Program Committee (JPC)—consisting of three Ghanaians representing academia, policymakers, care providers, and NGOs, and three Dutch scientists with backgrounds in health, biomedical, and social sciences—was constituted to perform such duties as policy decision making and awarding grants. A review of this program by Enyimayew et al. (2006) concluded that the research agenda was consistent with Ghana’s health sector priorities. The review also concluded that collaboration enhanced local research capacity, although the number of trained researchers was considered inadequate for conducting research to inform policy and to improve health service delivery. Beginning in 2009, the Doris Duke Charitable Foundation’s Africa Initiative funded research capacity training in five countries, including Ghana. Funding was used to support infrastructure, research training, and mentoring. Sustainability at the end of such project funding is often problematic, and there are no internationally shared metrics for measuring such a project’s impact (Hedt-Gauthier et al. 2017). Although systematic evaluation of the impact of these initiatives on development of research ethics capacity is rare, one assumes that these consortia and activities require high-quality ethical review and oversight as well as specific efforts to translate valid and reproducible research findings into improved health. They also pool resources to produce the science and ensure its integrity, which may be unaffordable to individual institutions. Innovative practice in research ethics is also evident in Ghana. For example, Tindana and colleagues developed a culturally appropriate framework for concerns about broad consent in H3Africa (Tindana et al. 2019). Facilitation of indigenous ethics for genomic studies is important. Entrustment, Tindana et al. argue, requires establishing responsibilities to use study samples widely and reciprocating by providing tangible health benefits. Research institutions with oversight from RECs should remain the stewards of these samples, attending to both their scientific and their cultural value. Entrustment requires a culture of RI, clear and transparent institutional policies, and RECs actively monitoring research to protect participants from the possible consequences of misplaced trust (Tindana et al. 2019). In a setting where these ingredients are absent, or where cultural values are at variance with entrustment, such a model may not promote RI. Exploring the problematic translation of bioethics between the Global North and South and between resource-rich and resource-poor countries, Miles and Laar (2018) argue that for such standards to be applicable globally, they must more directly engage the dialectical tension arising between cultural diversity and communitarianism. It is often said that western bioethics is excessively individualistic and that Africa, for example, is more communitarian (Beauchamp and Childress 2001; Coleman 2017; Gbadegesin 1993). These observations should serve as a reminder that standards of research quality may not be universal. Research Regulatory Structure and the Ethics Apparatus in Ghana Ghana has had a long history of health research–related engagements internationally and of implementation at the national and sub-national levels. Ghana has drawn and continues to draw on existing international guidelines and local knowledge to shape the development of its health research architecture, health research, and health service delivery protocols. This section describes Ghana’s health and public sector institutions as capacitated, tooled, and resourced to oversee, manage, and conduct health service research. Landmark projects and research efforts that emblematize Ghana’s post-independence development of health research systems include the Danfa Project in 1970, the Brong Ahafo Rural Integrated Development Project (Adjei and Gyapong 1999), and the establishment of the Noguchi Memorial Institute for Medical Research (NMIMR) in 1979 at the University of Ghana. The NMIMR was envisioned to be a world-class institute capable of conducting high-quality, cutting-edge research and training in the biomedical sciences. It is. Indeed, the NMIMR would become home to Ghana’s first IRB in 2000. However, Seddoh et al. (2015) note that the 1982 Report of the Council for Health Research and Development (COHRED) is most inspirational in charting the future course of health research in Ghana. Following the COHRED report, in 1992 Ghana developed a framework of Essential National Health Research that led to the development of five-year policy framework on health research development (Ministry of Health 1992). Seddoh et al. (2015) note that document set out the mechanisms for establishing a research agenda for the Ghana Health Sector, including the mechanisms for capacity development and for coordination of research. This eventually led to the establishment of the national Health Research Unit (HRU) in the Ministry of Health in 1994. In addition, three field research centers—Navrongo, Dodowa, and Kintampo Health Research Centers— were planned and established. Seddoh et al.’s work shows researchers’ high awareness of international ethics policies supporting health research. Of note, there is currently no law on health research in Ghana, although the Ministry of Health is generally acknowledged as having de facto responsibility for coordinating and providing leadership for setting national research priorities standards and regulating conduct of research countrywide (Seddoh et al. 2015). The Ghana Food and Drugs Authority (FDA), an agency of the Ministry, was established in August 1997 with the national regulatory authority to regulate food, drugs, and other products, as well as to provide guidelines for the authorization of clinical trials in Ghana involving medicines, food supplements, vaccines, and medical devices. The FDA also provides Ghanaian investigators with Guidelines for Good Clinical Practice for conducting clinical trials. However, the FDA does not provide ethical or legal standards for research not involving clinical trials: e.g., epidemiological research, genomics, public health interventions, health informatics, and other health sciences. Nor does the FDA offer a nation-wide definition of what constitutes RM. As described earlier, yet another important RI infrastructure in Ghana is Ghana Health Service (GHS), which administers government health services nationwide. GHS has an Ethics and Research Management Department to ensure the development of quality and consistency in all types of research conducted within the GHS. The Department hosts the Ghana Health Service Ethical Review Committee (GHS ERC), which reviews and approves all proposals for research to be conducted in GHS facilities or by GHS personnel. Aside from the GHS ERC, three other Health Research Centers of the GHS each have independent Research Ethics Committees (RECs). Each has a Federal-Wide Assurance (FWA) agreement with the United States Office of Human Research Protections (OHRP), but not with the US Office of Research Integrity (ORI). Of note, the governance of clinical trials is placed under the FDA in line with the dictates of the Public Health Act, 2012 - Act. No. 851 (Republic of Ghana 2012), Part 7 of which defines the requirements for undertaking clinical trials and how this might be regulated. In the absence of a national coordinating body, primary responsibility for RI and preventing RM lies within institutions. Ghana’s Council for Scientific and Industrial Research (CSIR) has the mandate to coordinate research activities in the country. The CSIR has an in-house ethics committee that support institutions requiring their review and ethical comment in support of grant applications. The University of Ghana (UG) offers an excellent example of an institution that has sought to develop its own research integrity infrastructure. In the last decade, UG has taken several steps to develop their institutional capacities related to RI. There are currently five IRBs at UG: (1) The Noguchi Memorial Institute for Medical Research—Institutional Review Board, including (2) an active Institutional Animal Care and Use Committee, (3) The Ethical and Protocol Review Committee of the College of Health Sciences, Korle-Bu Campus, (4) The Ethics Committee for the Humanities, and (5) The Ethics Committee for Basic and Applied Sciences. Additionally, UG’s Office of Research, Innovation, and Development (ORID) was established in 2010, creating a central office to promote, coordinate, and facilitate research activities in the university. ORID also leads the development of UG’s strategic plans, including its business plan and fundraising strategies. Over the last half decade, ORID has facilitated the development of various research policies, including (1) research policy, (2) research ethics policy, (3) intellectual property policy, and (4) research misconduct policy. ORID has recently created a Research Integrity Unit (RIU), which serves as UG’s independent hub for handling issues regarding and relating to RM. Other universities, teaching hospitals, and the Christian Health Agency, Ghana, contribute to the 17 research ethics committees currently in Ghana. An Association, the Ghana Administrators of Research Ethics Committees (GHAAREC) now exists, providing a platform for networking and sensitizing the public and research stakeholders about the need for proper research ethics as well as promoting the rights, welfare, and safety of research participants through efficient and effective research ethics administration. The guidance provided by the CSIR, FDA, GHS, UG, and the many RECs in Ghana make up a strong foundation to promote ethical conduct within biomedical research. However, gaps remain. Staffing RECs, establishing a national ethical framework for non-clinical research endeavors, assisting universities and other institutions in their oversight roles, developing RI training capacities, and attaining FWAs from OHRP and ORI within Ghana are important issues. To provide the needed facilitation in addressing some of the issues delineated above, the ethics apparatus in Ghana needs a National Research Ethics Council (NREC). The Council would, amongst others, set national norms and standards for conducting research, registering, and auditing RECs, adjudicate complaints about the functioning of RECs/IRBs, determine guidelines for the functioning of RECs, and promulgate regulations on RI. None of these currently exists. Some of the outlined issues will be addressed by two projects: first, the Ghanaian Research Integrity Development (GRID) Project; second, the NYU-UG Research Integrity Training Program (NYUUG RITP). These efforts are currently jointly being implemented by NYU School of Medicine and the UG School of Public Health. The latter effort aims to build sustainable research ethics and integrity capacity in Ghana. It will also establish a Bioethics Program, which will provide master’s degrees in Bioethics at the University of Ghana, the first program of its kind in the country (Caplan et al. 2018). The Project—Ghanaian Research Integrity Development (GRID) To date the major focus of the extensive effort to build research capacity in LMICs has been the establishment of ERCs/RECs/IRBs. Very little work has addressed the broader construct of RI, which incorporates the scientific quality of research, including prevention of RM and preservation of a valid scientific record in addition to ethical engagement with research participants and the public. A review of health research capacity development concludes that the most effective approach is locally built and led. Local groups and individuals often have a strong understanding of evidence gaps and of the political and cultural context (Franzen et al. 2017). Also important are the social and political structures in each country that support production of research with integrity (Mormina 2018). Motivated by this, researchers at NYU and in Ghana, with funding from ORI, implemented a project titled “Ghanaian Research Integrity Development (GRID)”. The project aimed to describe the relevant policies supporting RI in Ghana, determine current challenges to RI through self-assessment, and to formulate a plan to address them through RI capacity building. GRID engaged a group of sixty Ghanaian researchers, institutional and governmental officials, and was facilitated by local and NYU-based participants. GRID’s planning committee generated a list of relevant RI topics, which were prioritized for inclusion at a national conference. The logic model for this project is based on the notion of a complex adaptive system—a network of local and international institutions, scientists, regulators, national strategy, and health care system users of research, each with expectations and incentives that affect the production and use of science. The model is: (1) Assess participant perceptions of the level of RI, RM, and contributing conditions, informed in part by policy and position documents. This approach is congruent with current methods cited earlier since data about whistleblower allegations and disposition of RM cases are considered to be confidential and thus not accessible for study. (2) Since RI and prevention and detection of RM require some level of self-governance by the scientific community, this body should analyze perceptions, policies, and whatever empirical evidence exists, and determine its own judgment about adequacy of current practice and policy, and identify areas in need of improvement. (3) Develop a plan for desired individual, institutional, and national-level changes, including additional direct measures of areas of concern. Working through this process should create learning and accountability, which must be made explicit and based as much as possible on consensus. A survey serving as a needs-assessment tool was initially developed by Sergio Litewka and Elizabeth Heitman (Litewka and Heitman 2017) and modified for use in Ghana (Step 1 in the logic model), asking about institutional mechanisms, significant issues, and availability of formal RI educational activities. The survey aimed to capture, from the vantage point of researchers and administrators in Ghana, the current landscape of institutional approaches to research integrity and how those approaches might be improved. We invited the sixty researchers and administrators who registered for the GRID conference to participate in the survey. Our aim was to capture their perceptions of RI-related issues, the mechanisms their institutions have for promoting RI, the importance their institutions assign to these matters, and the challenges they presently face in those efforts. We specifically probed participants about the following issues: fabrication and falsification of data; plagiarism; data management, ownership, and sharing; financial COIs; misattributed or disputed authorship; and retaliation or fear of retaliation against reporters of RM. The questionnaire was anonymous and administered online using REDCap prior to the conference. Thirty responded. Our survey protocol obtained a priori ethics approval from the GHS ERC (Approval # GHS-ERC011/06/18). Data from the survey revealed that most participants perceived RI issues as significant problems in their institutions. For some of the issues, such as authorship, most participants reported there were no institutional mechanisms or formal educational opportunities to address them. For other issues, such as plagiarism, these perceptions occurred despite the fact that participants’ institutions had defined mechanisms or formal educational or training opportunities for addressing those issues. These results, which are the findings of self-assessment, suggest that there is work to be done to promote RI in Ghanaian institutions. A more fine-grained understanding of what that work must involve will emerge from our future research projects. The NYU‑UG Research Integrity Training Program (NYU‑UG RITP) The NYU-UG Research Integrity Training Program (R25-TW-010886) (Caplan et al. 2018) aims to address the gaps revealed by GRID. The program is a collaborative endeavor between NYU and UG, and is funded by the Fogarty International Center, U.S. National Institutes of Health. It has developed a fellowship program in research integrity to create a critical mass of faculty and other leaders with expertise in research ethics, RI, and research governance in Ghana. Its thirty fellows, divided in cohorts of ten over three years, take courses with faculty from both institutions in the history and philosophy of research ethics, research integrity, and developing collaborative research outputs. The fellows represent institutions across Ghana and are diverse in their professional backgrounds: they have professional or graduate degrees in medicine, nursing, social sciences, philosophy, law, and health sciences. Upon their completion of the fellowship, six fellows will matriculate in the master’s degree program in bioethics at the NYU Center for Bioethics. Beyond this stage of training, the ultimate goal of the project is to establish by its fifth year a Bioethics Program at the UG School of Public Health leading to master’s degrees in Bioethics. NYU-UG RITP is informed and animated by the belief that building capacity the local production of knowledge requires capacity building in research ethics and RI. It aims to meet the RI- and RM-related challenges and opportunities described above by training current practitioners and future leaders who will reshape the science culture and impact institutions in ways that promote RI and prevent RM. Through this program, Ghanaian researchers are merging world-class science with world-class ethics. Discussion In this paper, we discuss the state of the health research climate and research integrity in Ghana. We focus broadly on policies, structures and capabilities in Ghana, and conclude with recent efforts by NYU, UG, and other partners to develop research integrity in Ghana. As in other LMICs, private multinational organizations and foreign governments fund most of the health research in Ghana, raising unique issues related to RI. While there are some mechanisms in place to promote RI and prevent RM, Ghana is one of many countries in which primary responsibility for RI and prevention of RM lies within institutions, given the absence of a national coordinating body. Over the years, emphasis has been put on establishing IRBs/RECs. Currently, Ghanaian universities, research centers, teaching hospitals, governmental agencies, and quasi-governmental agencies contribute to a total of 18 IRBs/RECs. Little to no effort is geared toward RI. This is not unique to Ghana. The significant efforts that have been invested in building capacity for human subjects protection in LMIC settings (Glickman et al. 2009; Hyder and Rattani 2014; Nchinda 2002; Petryna 2009; Strosberg et al. 2013; WHO 1996) have not addressed the broader construct of RI (Ana et al. 2013). There is absence of evidence to show that the many and growing Ghanaian IRBs effectively carry out their oversight or RI-promoting responsibilities. A recent empirical review of IRBs in Sub-Saharan Africa concluded that current challenges to IRBs include a “lack of membership diversity, scarcity of resources, insufficient training of members, inadequate capacity to review and monitor studies, and lack of national ethics guidelines and accreditation” (Silaigwana and Wassenaar 2015). Others have noted that IRB members who lack adequate training are more likely to approve ethically risky research, especially given the motivation to impress funders and maintain international collaborative relationships (Ndebele et al. 2014). While comparable assessment has been done of the Ghana IRBs, our collective appreciation of their circumstances makes us believe that they do not fare differently. Indeed, the effort made by Seddoh et al. (2015) to map health research institutions in Ghana identified other weaknesses of the Ghanaian health research apparatus. They made several recommendations, including improving the quality of research, RI, and the general health research landscape in Ghana. On the basis of these, we argue that the structures for detecting and preventing of RM and improving RI are less well developed in Ghana. Other actions recommended by Seddoh et al. (2015) were the development of a national health research priorities and strategic plan, and continued capacity building for research staff. In these recommendations, we identify opportunities for improving RI in Ghana, but also challenges. A deliberate inclusion of RI in the national research priorities and strategic plan is an opportunity. Also equipping researchers with required RI competencies through higher education capacity-building initiatives and health research management training is another opportunity. However, it is worth noting that capacity building, though an opportunity, is also fraught with challenges. Laar and colleagues share their experiences and challenges faced in their efforts to strengthen nutrition capacity in Africa (Laar et al. 2017). They note that, for such capacity-building initiatives to have sustainable impact, a cocktail comprising higher education initiatives, shortterm training, experiential learning, and continual/lateral mentoring is required. In the current context, such capacity-building initiatives would include structured RI or responsible conduct of research (RCR) trainings at the individual or at institutional levels. RCR as a concept covers core norms, principles, regulations, and rules governing the practice of research. The history, purpose, and future of instruction in RCR as well as what mentoring and training in RCR have been addressed elsewhere (Anderson et al. 2007; Steneck and Bulger 2007). Satalkar and Shaw (2019) discuss the significance of structured RI or RCR trainings at institutional levels, noting that such training will strengthen faith in RI among those who value honesty, transparency, and trustworthiness in their work. For those who are at risk of compromising RI if the external pressure to succeed gets too high, RCR training for such individuals, they note, could “strengthen the angel sitting on one’s shoulder whereas institutional structures, rigorous supervision and internal peer review will keep ‘the devil’ sitting on the other shoulder in check” (Satalkar and Shaw 2019, p. 10). Many calls exist for the establishment of an independent National Health Research and Ethics Council backed by legislature (Seddoh et al. 2015). Its proposed mandate includes: setting national norms and standards for conducting research; developing guidelines for the functioning of research ethics committees/IRBs; registering and auditing RECs; publishing a national registry of credible research institutions and organizations doing research in Ghana. As an authoritative national regulatory body that governs the conduct of research, the Council’s RI-promoting duties should also include adjudicating on RI- and RM-related disagreements and complaints. While there is growing base of research ethics infrastructure, as described earlier, both individual capacities and institutional RI structures are weak. There is an urgent need to address these. Our RI training initiatives—the GRID and NYU-UG RITP, pay heed. These projects have facilitated the formation of the Ghana Research Ethics Consortium (GREC), which among others, would oversee the organization of the Annual GRID Conference. With representation from all public universities in Ghana, the GHS, CSIR, FDA, Ghana Standards Authority, four Teaching Hospitals, GHAAREC, private Universities and other research institutions, the is well placed to provide strategic guidance on RI in Ghana—in line with Mitcham’s notion of coresponsibility for RI (Mitcham 2003). Limitations The results of the online survey cannot be generalized although they provide valuable insight into the subject of RI. Also, the respondents of our online survey self selected themselves and hence the results need to be interpreted keeping self-selection bias in mind. Those who were particularly interested in the topic of research integrity are more likely to have responded to our call for research participation. In spite of our attempts to include researchers across the three major ecological zones of Ghana, the participation by researchers from the Savanah and Coastal was minimal.

Within the curriculum of graduate health care education, a course in research is typically included as the foundation for the student’s future role in research. The research course is designed to build on undergraduate work in research analysis and focuses on building graduate level skills in the methodological components of research. These skills include design, sampling, data collection, data management and analysis, research application, and ethical considerations specific to health care research. Completion of the online and easily accessible Collaborative Institutional Training Initiative (CITI) Human Subject Research program1 is a required element of our graduate research course. However, faculty noted gaps in the students’ interpretation of research articles and research proposals, demonstrating a lack of analytical and reasoning skills in the application of research ethics. Although basic research ethics knowledge was acquired through CITI training, the analytical and reasoning skills of human subjects’ protection in planned research practice were inconsistent. This lack of proficiency in analyzing benefits and risks, identifying elements required for consent, protecting patient identity, and evaluating the scientific basis for the study indicated a deficit in students’ ability to apply the knowledge regarding the protection of human research subjects. Purpose The purpose of this study was to investigate the effect of educational interventions on the research ethics analytical and reasoning skills of 4 classes of graduate health care students over a 2-year period. The study evaluated the effect of an online CITI Human Subjects program and an in-person mock institutional review board (IRB), with students serving as IRB committee reviewers of submissions containing errors and ethical issues. Review of the Literature After the disclosure of a history of gross ethical misconduct in research, the United States enacted the National Research Act of 1974.2 This legislation created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The charge of the Commission was to summarize the basic ethical principles foundational to the conduct of ethical research involving human subjects and establish guidelines that confirm that human subject research is conducted in accordance with those principles. The Belmont Report is the document that summarizes the basic ethical principles of autonomy, beneficence, and justice.2 The correct interpretation and implementation of these guidelines is an important consideration in this era of exploding clinical research. A frequent requirement across graduate health care research courses is the completion of the online self-directed University of Miami’s CITI program on human subjects. The course is interspersed with testing on the regulatory and ethical issues addressed and the learner’s ability to memorize factual information related to basic ethics knowledge.1 Braunschweiger and Goodman3 reported that CITI instruction increased basic knowledge in human subject protection issues, improved users’ confidence to ethically perform research, and documented compliance with federal regulations regarding human subjects protections. Educating graduate nursing students in the application of research ethics is an essential competency because nurses are prepared to address gaps resulting from growing health care needs, use evidence-based practice, contribute to health care services delivery, and function on interprofessional teams.4 Graduate level nurses should be prepared to apply research outcomeswithin the clinical practice setting, resolve practice problems, and disseminate results within the setting and in wider venues to advance nursing practice.5 Although the importance of graduate nurses achieving this competency is evident in professional standards, the literature does not demonstrate its consistent use in practice. In an international survey conducted with 1675 nursing journal reviewers, Broome et al6 found that approximately 20% of the reviewers reported various ethical dilemmas that included insufficient protections of human subjects such as the possible identification of participants, failure to obtain IRB approval, and conflicts of interest. Additional concerns were the use of questionable enrollment processes, procedures of obtaining informed consent, and lack of clarity for protection requirements globally.6 Mitchell and Carroll7 reported the lack of ethical research knowledge in their study of doctoral nursing students. The authors used a case study approach to investigate the importance of student engagement with the regulatory process, identify problems of knowledge or understanding transfer, increase awareness of culturally specific issues that might occur during the conduct of a study, and understand academic theft. Even after students became aware of research regulatory guidelines, students were unable to identify needed researcherinitiated changes to the research protocolwithout involvement of the research ethics committee. Students did not recognize the failure to conduct an adequate search of existing research before beginning a new project as ethical issues.7 A recent study by Cho and Shin8 used a blended approach that first addressed knowledge in the online arena followed by classroom experience using case studies. In their study, the authors validated the lack of published research addressing research ethics education and discussed their students’ lack of understanding of the IRB process and limited knowledge of publication ethics. Content and strategies suggested to teach research ethics address ethical principles, the context of the research, and provision of practice opportunities. Sherrod et al9 stressed that students be taught the basic principles of research using examples from research studies that require IRB approval because IRBs are responsible to ensure ethical treatment of research subjects. Sweet10 used a mock IRB simulation with hypothetical research projects among undergraduate sociology students to build their skills in ethical decision-making. As a consequence of the mock IRB activities, students increased their ability to evaluate studies on the basis of ethical merit and empirical rigor. Harrowing et al11 suggested that in addition to focusing on the principles of respect for persons, justice, and beneficence the social and cultural context of the setting should be included in the ethical assessment process. The professional literature provides little evidence-based guidance on effective strategies to promote the development of ethical research analytical and reasoning skills for graduate health care students. Although the literature suggests graduate students first learn and then have the opportunity to apply ethical research principles, the question remains whether these skills can be acquired through an online course or if an interactive in-person simulation will produce a greater increase in the needed analytical and reasoning skills. Method Design This study used a quasi-experimental pretest-posttest design to study the effect of 2 educational interventions in the form of (1) online University of Miami’s CITI Human Research Subject program and (2) an educational session on the analytic and reasoning skills related to research ethics, including an in-person mock IRB experience. The study was approved by the university’s IRB as an exempt study. Sample The study population consisted of students enrolled in the 4 graduate research courses conducted on the university’smain campus in 2015 and 2016. The students included graduate nursing and nutrition students. Four classes totaling 64 students were invited to participate as volunteers in the full study. Two students declined to participate, resulting in a total sample of 62 students. Educational Interventions The first educational intervention, the CITI Human Research program, consisted of 18 online modules with tests after each module. Each student was required to submit the certificate indicating successful completion of the modules. The second educational intervention was the mock IRB session. The mock IRB session included student prework, class discussion, and a mock IRB. The prework included review of a faculty-narrated slide presentation of IRB structure, function, roles, and responsibilities; Title 45 Code of Federal Regulations12; Requirements for the Ethical and Regulatory Review of Research by National Institutes of Health (NIH) IRBs13; and university IRB protocol submission forms. Class discussion covered several ethical issues such as informed consent, financial reward, confidentiality, authorship, and a summary of the HIPAA of 1996.14 The mock IRB was set up as a standard IRB committee meetingwith 5 teams of students reviewing, evaluating, and presenting recommendations that addressed human subject ethical research deficiencies for their specific case. Instruments The data were collected using the Research Ethics Knowledge and Skills Assessment (REKASA) to assess analytic and reasoning skills important to research ethics.15 The REKASA is a 41-item assessment with established face and content validity and a Cronbach’s ! of .837.15 The REKASA included 22 multiple choice questions, 2 true-false questions, 9 short answer questions, and 8 questions based on 2 case studies, for a total of 41 questions. An additional 10 demographic questions were added to the pretest. Five previously approved full and expedited IRB applications were altered by course faculty, deleting/adding information to highlight ethical issues in informed consent, vulnerable populations, risk assessment, conflict of interest, scientific issues, confidentiality, and ethical issues beyond IRB authority. Following university guidelines, the applications were placed on university forms to simulate the actual IRB committee environment. Two experienced research faculty reviewed the IRB applications for clarity and for face and content validity based on their experience with IRB applications and ethical issues in research. Procedure Researchers without responsibility for assigning grades for the subjects were responsible for obtaining informed consent and administering pretest/posttests so that potential subjects would not feel coerced or obligated to participate. The REKASA was given as the pretest immediately after obtaining consent, as posttest-1 after the CITI course was completed, and as posttest-2 after the mock IRB. Each time the instrument was given, the questions were reordered. Each student had 60 minutes to complete the pen-and-paper REKASA at each administration. The students were assigned alphabetically in groups of 3 to 4 to serve as IRB committee members and present their group review of a faculty-altered IRB proposal. The university IRB proposal review template was used to provide structure for their review. At the mock IRB meeting, each group presented their assigned protocol, and the entire class discussed the group’s assessment and recommendations to approve, modify, or disapprove the proposal. Data Analysis The REKASA was scored by recording 1 point per item for each correct response to obtain a total score for each subject and test. Multiple-choice and true-false items were scored and entered into Microsoft Excel (Microsoft, Redmond, Washington). All short answer and case study questions were scored by 2 researchers and reconciled before entering the scores into Excel. The data were then analyzed using IBM SPSS version 21 (Armonk, NY) to conduct a repeated measures analysis of variance (ANOVA). A total of 62 participants completed the pretest; 50 participants completed posttest-1 after the CITI training, and 42 participants completed posttest-2 after the education session with the interactive mock IRB review. The final sample consisted of the 37 participants who completed all 3 tests and completed both the CITI training and themock IRB session. Participant drop-out rate of 40.3% is related to withdrawal from class, the researchers’ decision not to administer the REKASA for participants who missed class or arrived more than 15 minutes late to class, and those who voluntarily withdrew from the study. Results The subjects participating in the study included nursing students (n = 31, 83.8%) and nutrition students (n = 6, 16.2%). Most participants (n = 36, 97.3%) had a bachelor’s degree as their highest education level. Participation in research education was closely divided with most subjects (n = 21, 56.8%), having attended a previous research course. Previous IRB experience was limited with most students (n = 36, 97.3%), having no experience with an IRB. Of the 37 participants, most (n = 30, 81.1%) had never completed CITI or NIH Protecting Human Research Participants Training.16 The subjectswere closely divided between ages 20 to 33 years (n = 19, 51%) and ages 34 years and older (n = 18, 48.6%). Of the 31 nurses in the study, most were recent graduates, having graduated from nursing school less than 5 years earlier (n = 19, 61.3%). Subjects were predominately female (n = 31, 83.8%), Hispanic (n = 22, 59.5%), and selfidentified as white (n = 32, 86.5%). The remaining subjects identified racially as black (n = 2, 5.4%), Asian (n = 1, 2.7%), and other (n = 2, 5.4%). The work experience of the participants was primarily acute care nursing (n = 22, 59.5%), non acute care nursing (n = 7, 18.9%), non-nursing (n = 5, 13.5%), and no full-time work experience (n = 3, 8.1%). The REKASA provided outcome measurements for students’ acquisition of research analytic and reasoning skills important to research ethics through a baseline pretest before education, posttest-1 after the CITI education intervention, and posttest-2 after the mock IRB education session. The mean scores for the REKASA showed a pretest mean of 26.27 (SD, 5.00), post-CITI mean of 29.49 (SD, 4.28), and a post– mock IRB mean of 31.46 (SD, 5.55). A repeated measures ANOVA with a Huynh-Feldt correctiondemonstrated thatmeanknowledge and skills differed significantly between the education methods of no training (baseline), CITI training (posttest-1), and mock IRB (posttest- 2) (F [2, 72] = 26.92, P G .001, )p 2 = .428). Posthoc tests using the Bonferroni correction showed that there were significant differences between each set of means including pretest and posttest-1 (j3.216, P G .001) and posttest-2 (j5.189, P G .001) and between the means of posttest-1 and posttest-2 (j1.973, P = .01), indicating a significant increase from pretest to posttest-1 and from posttest-1 to posttest-2. Therefore, we concluded that CITI training increased the research ethics knowledge and skills over baseline and that an education session including a mock IRB increased the research ethics knowledge and skills over baseline and again over CITI training. The )p 2 indicates that 42.8% of the variance inmean test scores measuring analytical and reasoning skills important to research ethics is accounted for by the educational interventions. At baseline, before educational intervention, most students (n = 32, 86%) demonstrated that they had an understanding of selected researcher responsibilities such as the safekeeping of collected data; honesty in science including data accuracy, conflicts of interest, and integrity; and subject protection including the risk of introducing new drugs, disclosure of risks and benefits in the patient’s own language, exclusion of subjects not meeting criteria, and subject’s right to decline to participate. These correct baseline scores correlated with the students’ ability to make ethical distinctions in clinically relevant areas versus areas of social justice. The highest number of correct responses reflected situations that were common in the clinical practice of the nurse such as an understanding of the subject’s risks in the informed consent process. At baseline, students varied in their understanding of the Belmont principles with all of the students correctly interpreting the ethical principle of beneficence and a minority correctly interpreting autonomy (n = 12, 32%) and justice (n = 9, 24%). After CITI, interpretation of justice increased (n = 14, 38%) and, after mock IRB, increased further (n = 22, 60%). Similar increases were noted for autonomy with increases after CITI (n = 21, 54%) and mock IRB (n = 27, 73%). From baseline to posttest-2, the greatest increase in students’ correct responses, more than 40%, in research ethics analytical and reasoning skills, included methods to reduce bias, elements of an ethical design, and the Belmont principles of autonomy and justice. Students’ correct answers increased from baseline to posttest-2 at least 30% related to understanding of ethical principles, recognition of moral concerns, tension between 2 principles, and actions/changes to address ethical principles. These increases were also seen in understanding the work of the IRB. An area that did not showsubstantive increase in understanding was the balance of risk and benefits with most students being intolerant of any risk for the subject. The second area that did not show substantive increase was related to the amount of information to provide to every subject as a part of informed consent. A large portion of the students determined that subjects should automatically be provided the full research protocol at the time of signing the consent. Discussion Before this educational intervention, most students had exposure to ethical decisions in their clinical practice or education contributing to slightly more than a quarter of the 41 questions (27%) being answered correctly by most students (n = 32, 86%). It also indicates the need for substantive graduate research ethics education. This study, the first of its design, demonstrates the effectiveness of combining 2 educational approaches for the ethical research education of graduate health care students. The study results support the efficacy of CITI training3 for increasing knowledge and further demonstrates the improvement possible with the addition of a faceto- face component, which includes a mock IRB. Anecdotally, positive results were noted in an undergraduate research sociology class using a mock IRB conducted by Sweet10; however, a mock IRB experience is not reported in scholarly literature as being an educational strategy for graduate health care students. The study’s approach expanded on Sherrod and colleagues’9 case study recommendation by adding the mock IRB simulation experience. The American Association of Colleges of Nursing’s call for graduate nursing competency in research development and usage necessitates nurses understanding risk-benefit analysis and the ethical principles of beneficence, autonomy, and justice.4,5 The goal of students in this study was to ensure subject benefits and removal of all risk. At baseline, the first interest of these students was the health and well-being of the individual client, reflecting the patient advocate role in the clinical setting. Graduate education enables students to consider that new medications and treatments will not be available without research in which risks may be present and benefit not guaranteed. This study’s results support the use of a mock IRB simulation to enhance skill in research ethics decision-making. An area that did not show substantive increasewas related to the amount of information to provide to every subject as a part of informed consent. For example, 40% of the students determined that subjects should automatically be provided the full research protocol at the time of signing the consent. This belief in patient’s autonomy is not balanced with an understanding of a patient’s health care literacy. In a highly ethnically diverse community, the application of ethical principles is difficult to envision without situational context. As Harrowing et al11 stated, application of ethics needs to be considered in the context of the participants’ lived experiences. Providing the mock IRB experience enabled students to place research within the context of dilemmas research subjects potentially face. Limitations Limitations to the study included the limited sample size, single site, test length, and participant burden. Despite these limitations, the effect size indicated that the educational interventions accounted for 42.8% of the variance in the ethical analytical and reasoning skills. Future studies should consider offering testing at alternate times to accommodate late-arriving students. The length of the test/time restriction was a barrier because students did not consistently complete the short answer questions. Conclusion This study supports the use of CITI Human Subjects Research training to increase research ethics knowledge. After the completion of CITI training, the addition of a mock IRB simulation significantly increases students’ research ethics analytical and reasoning skills. The future role of health care graduates in clinical research requires faculty to explore evidence-based methods that promote acquisition of this essential competency. Ongoing research using multiple education strategies will contribute to the evidence-based approaches available to educators as they work with graduate health care students.

The current research climate in UKhigher education institutions (HEIs) is one where research assessment determines amajor part of institutional funding. Bidding for competitive funding takes place in an increasingly competitive environment, and research impact is taking an ever more important role in the evaluation of research outcomes. Institutions need to be fully informed about research being undertaken within their walls, and organisational units are expected to provide rapid responses to requests for information about research activities. Such demands give rise to a need for easy access to up to date and reliable data. Universities often posses the data they need: the problem lies in finding appropriate data, accessing the data, aggregating data into a useful form and providing it in a useful output format. The data may be dispersed across units, in a format that makes it inaccessible, and need considerable manual work to make it useful. When considering the prospect of raising income, rankings and the institution’s image, universities are keen not to be left behind their competitors. Coupled with this there is a desire to cut costs and improve efficiency as budgets are squeezed, a situation not helped by inefficient data management, for example unacceptable duplication of effort when creating data. The picture is further coloured by the need to attract high quality research students and staff by publicizing the institution’s research. 1.1. Research activity data—a sub-set of data Management of research at an HEI requires data gathered from a wide array of sources including financial, human resources and research services. This article and the project described focuses on a sub-set of that data that we are calling ‘research activity data.’ Research activity data is data about research and researchers that describes what research is taking or has taken place, who is involved, who is funding it and what the outputs are. It is not concerned with research processes such as managing a research project budget. 1.1.1. Research activity data in the institutional repository Like many HEIs, the University of Oxford has implemented and is continuing to develop an institutional repository for research materials. The Oxford University Research Archive (ORA) (ORA, 2009) is a growing collection of research materials produced by scholars at the university. The repository team at Oxford wishes to encourage use of the repository by making deposit easy and, in an ideal world, barely noticeable to authors and creators. This might include bulk uploads of metadata and content from other sources, automatic metadata generation and deposit direct from a personal workspace. The large amount of material produced by researchers coupled with intense pressures on academics’ time means that they are generally unwilling to fill in a form each time they wish to deposit an item, especially when that form requests data which they have supplied endless numbers of times to different people. It is therefore sensible to derive metadata for items in ORA from existing data stores wherever possible. However, it proved difficult to identify and obtain data for re-use in this way. The technical architecture of the Oxford repository has a modular design and digital object structure. In drawing up this design, there were plans for a registry to store data used to describe items in ORA, including that drawn from other sources. Data about people, departments and so on could be held as data entities within the registry. Each entity could be fully described and linked to other related entities. The team soon realised that the solution they were proposing had the potential to provide a number of additional benefits for the university. The emergence of this solution and the fortuitous simultaneous call for bids for funding issued by the JISC in their circular 7/08 Institutional Innovation Programme, prompted the team to propose a formal project. The resulting project, Building the Research Information Infrastructure (BRII) runs from September 2008 to March 2010. The aim of the project is the efficient sharing of research activity information using semantic web technologies. Data is harvested into the pilot registry, created as the main output of the project. Ontologies and taxonomies define and describe data objects (for example people, research groups, funding agencies, publications and research ‘themes’) to forge connections between them and provide web-based services to disseminate and re-use this information in new contexts. Data sources include disparate groups, both academic and administrative, from across the collegiate university. 1.1.2. Fit with existing and planned systems In large organisations management systems tend to deal with different aspects of research such as recording details of researchers or successful projects, and have some data in common with other systems. The ideal situation is where all systems fully integrate and interoperate with each other. In practice, this is not always the case. It is important that anynewsystem does not duplicate another existing (or planned) system, although theremaybe overlap in their operations. The BRII registry does not duplicate any current system, but provides a ‘view’ on existing data. It bridges a gap in the existing research information systems at Oxford, using an innovative design that offers additional benefits for integration and data sharing. Its semantic web design means that it does not have to interoperate technically with other systems. However, difficulties of obtaining data from some sources should not be underestimated. One area under scrutiny at a national level is the use of Common European Research Information Format (CERIF) (EuroCRIS, 2009) for research data exchange. Oxford is monitoring developments closely and participating in discussions about use of CERIF for the forthcoming research excellence framework (REF). The outcomes of such discussions and the ability to comply will be of vital interest to senior management and can be included in the future development of the BRII registry. Like others involved in this area, Oxford is waiting for the report from EXRI (2009) project (due December 2009) which is expected to result in recommendations for research data exchange formats for the UK. 1.2. External influences The need for access to research activity data is heavily influenced by the actions and requirements of bodies external to the university. The university cannot afford to ignore these requirements as money and prestige are at stake. The main external influence is probably the Higher Education Funding Council for England (HEFCE) and its forthcoming exercise, the REF, aimed at judging the quality of research (HEFCE, 2009). HEIs place great importance on their REF returns and the ability to gather the data and other information about the research taking place within their institutions is considered high priority. Funding agencies want evidence of how their money is being spent and judge the value for money that is being achieved. They therefore require recipients of research funding to report on their activities, outputs and outcomes. Compliance in so doing could affect decisions about funding of future research proposals by that team. Research activity data is used to create and inform publicity and communications materials in order to attract high quality staff and students. As such it plays an important role in defining the university’s place in the world, external perceptions of the institution, and the calibre of its research.Motivations 2.1. Levels of interest Early on in the project a stakeholder analysis of those involved with research activity information at the University of Oxford was undertaken. Interviews with stakeholders resulted in them being grouped by function: • Researchers: those who actively define and undertake research. • Administrative staff: those who manage research activities at different levels. • Strategists and disseminators: those whowork at university level and who are not directly involved with research activities. It is this third group with which this article is concerned. These individuals: ‘need to keep well informed so they can develop strategies for future research priorities and disseminate research information across the University, nationally or internationally. Roles within this category include heads of departments and divisions, research and strategy committees, vice chancellors and the Press Office. . .Strategists and disseminators do not produce research activity data but are heavy consumers of them. They are continuously fed with information which they have to interpret, adapt, reformat or publicise in different outlets. . .Products of their work could be classified as reports or summaries written in non-specialised language and whose scope covers all the University’s research disciplines. They possess a high level perspective of research activity at the University’ (Loureiro Koechlin, 2009). Academics’ activities are focussed on their specific subject area within their world of dispersed subject-related colleagues in which they operate. However, university level users comprising strategists and disseminators, work across the institution and must deal with all disciplines. It is this high level ‘university view’ where the need seems most pressing for research activity information. The information required by this group needs to be broad and general rather than at the level of detail needed by subject specialists. 2.2. Meeting the requirements of strategists and disseminators The interviews that formed the stakeholder analysis threw up a number of real-life questions from strategists and disseminators together with general areas of information need. 2.2.1. Problematic questions The types of questions posed by strategists and disseminators and requiring research activity information include: • What research is taking place in a particular research area (which may be cross- or multi-disciplinary research)? • Who is collaborating with whom (both internally and externally)? • What are the outputs associated with a particular project? • Who is working in a specific research area and collaborating with others in a specified country: a rapid response required? These sorts of queries do not require answers that give detailed information about the research and its findings: that is the domain of the research specialist. The importance of any system such as this for managers and disseminators lies in the ability to gather an overview of that research information and to be able to do so quickly. By contrast, the stakeholder analysis revealed that these high level needs are not of interest to the individual researcher. The questions posed above tend to arise in situations such as contact with existing or potential sources of funding, internal reporting, personal contact (such as a visit by influential individuals) or to identify research opportunities. Statutory and other reporting together with publicity are additional drivers for developing easy methods of gathering and reporting this information. 2.2.2. Current sources of data The top level university view of research activity often requires a general picture of activity across the university to be drawn. However, this overview can be elusive. It has to be built up from elements available in disparate and disconnected sources, a bit like doing a jigsaw comprising many separate pieces all stored in different rooms. One cause of this is that most current initiatives to collect research information within Oxford have their own aims and business processes at the forefront: making the data available for other purposes has until now been of secondary consideration or altogether out of scope. 2.2.3. Additional important factors As a general rule, the image of the institution as a whole is important to university level strategists and disseminators. Research activity data feeds publicity channels and can be used by external communications departments to provide information to a wide range of interested groups. This is particularly important when comparisons with other institutions are being drawn. Timing can be crucial. This might be because the immediate need to identify a specialist to field an interview with the press, or it might be because an urgent funding opportunity presents itself. In these sorts of scenarios, university level staff need information that can be obtained immediately and reliably and that is up to date. 3. Oxford research activity registry The main output of the BRII project is a pilot registry that has the potential to meet a number of the needs expressed by different stakeholders. It tackles the twomain challenges of research activity data: (i) provide data for the easy discovery of, access to and reuse of research activity information (for internal purposes) and (ii) to feed channels for the dissemination and promotion of Oxford research (the public facing aspect) as shown in Fig. 1. 3.1. Boundaries and scope The duration of the project (18 months) meant that clear boundaries had to be drawn as to what was possible within the timescale. It was decided that the project should focus on data that could be made publicly available and avoid the many difficulties associated with restricted data including access rights. For the purposes of the project, publicly available data is classed as that which could be made freely available via a web site. The registry can technically store restricted data. It remained important to respect the wishes of some groups not to participate because of the sensitivity of their data. 3.1.1. Data harvesting Data is harvested from existing sources causing minimal interference to those systems. It is important that data providers understand that the data in the registry mirrors the original source: ownership of the data is retained by the original data source. All data are marked with the source and any queries regarding accuracy or omissions are directed to the source. The aim is to harvest ‘live’ data. This is to ensure that data are as up to date and as accurate as possible. There are some instances where static data are acceptable such as a list of publications which will not change. Manual entry of data will be avoided. The ideal is to encourage systems whereby the data is ‘pushed’ into the registry. Some examples of systems capable of this have been discovered and can be used as examples of good practice for others. Types of data sources and their ease or harvesting and synchronization have been drawn up, as shown in Table 1. It is expected that the majority of data sources will fall into the fourth category, that which can be accessed, but does not have records of history or provenance, or obvious modification dates, and which will vary in their ease of inclusion. An assessment of the work needed to repeat data harvests from contributors beyond the life of the project and the most cost-effective developments to support export from key sources will be considered as a part of future development. Because data is being provided by contributors across the university, the registry is dependent on those sources for the quantity and quality of that data. This is an important point when considering the requirements of users of the data. Users want registry data to be comprehensive and accurate. But until there are many data contributors with comprehensive data to offer, the registry cannot be comprehensive. The project team managed this situation by stating clearly that the project will provide a pilot service to demonstrate the potential of the system. It was decided to focus on depth of coverage using a limited number of contributors rather than attempt to demonstrate breadth of coverage. Having said this, as many sources of content have been included as proved practical including from a selection of academic units, from at least one central administrative department and from one college. 3.1.2. Managing expectations There was a danger that the system would be perceived as a ‘cure-all,’ such was the desire for easy data access by senior managers. The team had to ensure that the users who were interviewed and expressed their requirements for obtaining data understood the types of data with which the project was concerned. For many, their needs for research information extended beyond that covered by the BRII registry. It was crucial that the project managed the expectations of the community. One means of achieving this was to document what the registry could be expected to offer by the project end and to publicise this information to key stakeholders. It was also important to describe the potential for future extension, additional functionalities and improvements based on wider explorations of user requirements, given adequate time and resources. 3.2. Additional benefits The implementation of the BRII registry is expected to result in a number of additional benefits to the institution and to users of the data, especially to strategists and disseminators. The data will not only be available to assist answering questions such as those listed above, it could also be re-used to save another individual re-creating that data as has previously been the case. This refers to the original problem that was identified with ORA and the desire not to have to re-enter data such as author name, department and so on. Having a single canonical source of the data not only aids efficiency by reducing duplication of effort when creating the data, it also encourages accuracy and consistency because only a single source has to be maintained, thereby eliminating the possibility of errors as the same data is duplicated. There are also efficiency gains in staff time as they are able to find and access the data they require quickly and easily. By making data more visible, frequently accessed and re-used, and by offering an easy to use error and omission reporting mechanism, data owners will be encouraged to improve the comprehensiveness and quality of their data. One of the main drivers for setting up Oxford’s digital repository was the long-term preservation of and continued access to the Library Services’ digital collections. Oxford’s digital asset management system (DAMS) which is being created and managed by Oxford University Library Services (OULS) has at its heart the storage and preservation of Oxford’s digital collections. The BRII registry is part of the DAMS and as such, the digital objects it comprises will be preserved along with Oxford’s other digital collections. This forms an important service for data owners and extra reassurance of the longevity of their data. 3.3. ‘Blue Pages’ and themed web site Although the development of the registry is the cornerstone of the BRII project, to demonstrate the possibilities for data re-use that it offers, two exemplar web applications are being constructed. The concept of the registry can be difficult to grasp: it is an invisible resource. Creating a visible manifestation of use of the data enables potential users to understand what might be possible and enables them to ‘see’ the data in action in a real-life situation. The two example uses of the data are the Oxford ‘Blue Pages’ and a themed web site. 3.3.1. Oxford ‘Blue Pages’ The Oxford ‘Blue Pages’ is an online directory of researchers and research activity at Oxford. It uses data held in the BRII registry to populate a freely available web site which enables users to search and browse for people, projects and other information about research at Oxford. Users can access a person’s profile page to find out which projects they are associated with, their affiliations (such as department and research group), their publications (including data drawn from ORA) and their collaborations. Other pages include similar ones for research projects. It is not unknown for members of Oxford University to have two or more personal web pages, plus information about their research activity held on other Oxford web sites. The ‘Blue Pages’ does not aim to replace existing personal webpages. Instead, it draws disparate data together making it accessible and searchable in one location. It also acts as a signpost, pointing the end user to the webpages or other sources that supplied data. The importance of the ‘Blue Pages’ for managers and disseminators lies in the fact that it can be used to find and access Oxford research activity data easily from a single location. The data can then be extracted for re-use as required. In fact, what happened was that the ‘Blue Pages’ started to become accepted as an important resource in their own right. The Blue Pages offers the opportunity to make connections between researchers and research activities. This is only possible when the user can obtain details across fields and departments. These connections might not have been previously apparent and therefore could be influential in identifying new research opportunities or collaborations. This type of information is crucial for strategists, particularly those involved with informing academic staff of funding opportunities. Extensive user testing was undertaken to assist with the design and functionality of the web interface. An iterative approach was taken to speed up development and to reduce wasted effort. Testers were selected from different communities across the university to ensure different viewpoints were taken into consideration. The quantity and quality of data available via the ‘Blue Pages’ is dependent on the data held in the registry and supplied by data contributors. This affects the comprehensiveness and accuracy of the data which understandably is of great importance to strategists and disseminators. However, these individuals are sometimes in a position to influence the maintenance and content of the original data sources. 3.3.2. Themed web site The data held in the BRII registry is being used to demonstrate an example of the API integration with a themed web site. A themed web site is a means to aggregate disparate data around a common selected theme. Users can then easily discover all activities included in the registry based around that theme, no matter where they are located. It is therefore possible to present a more subject-focused overview of research in areas such as cancer and neuroscience which cross a number of departments and cross-disciplinary topics such as climate change. The data can be combined with other data not held in the registry. The result might highlight potential collaborators and other connections across the university. For the purpose of demonstration the Medical Sciences Division will construct a web site integrating information about researchers across the Division, with information about opportunities for graduate studies. 3.4. Simplifying deposit in ORA Returning to the original reason for developing the BRII registry, one of its main functions is to simplify deposit in Oxford’s repository for digital research outputs, ORA. It will do this by offering auto completion of fields where the registry holds the data. There will be shared use of the subject descriptive thesauri held in the registry (for example the American Mathematical Society classification or author generated vocabularies created using software produced as an output of the BRII project). 4. Using semantic web and digital object model HEIs are generally large complex institutions. Oxford, it could be argued is more complex than most, with its devolved collegiate structure, nearly 120 libraries and around 7000 research active staff formed from a mixture of college, faculty and other appointments. Use of semantic web technologies coupled with the digital object model allows this complexity to be reflected within the system. Attempting to shoehorn the complex nature of the people, their associations and so on into a rigid database table structure would not only be difficult, it would limit the use of the system. Researchers can have many different roles, for example, a single individual might be a professor, a principal investigator on one project, a member of the research team on another and a head of department. All of these can be accommodated. Add to that membership of multiple departments and research institutes that draw people from many departments across academic divisions, and the depth of complexity begins to emerge. Within the registry there are a number of separate entities including ‘person’, ‘project’, ‘place’ and ‘funder’. These entities are defined and the relationships between them described using the RDF triple model. Each entity holds all the data about them including temporal data (see below). Some data is fixed (e.g. date of birth) and some more fluid (such as role of PI). In this way, importance is placed on the context of the data to ensure that source and provenance are recorded. The registry uses RDF ontologies and taxonomies to define and categorise data objects. This enables connections between researchers, grants and other entities to be forged. Wherever possible existing ontologies and taxonomies are used and, if necessary, extended. They include: • RES, a researcher profiling system developed in the Medical Sciences Division (MSD) (MSD, 2008). • The Academic Institution Internal Structure Ontology (AIISO) (Styles and Shabir, 2008). • The Participation Ontology (Styles et al., 2008) for describing the roles that people play (Styles and Wallace, 2008). • Academic Research Project Funding Ontology (ARPFO) a new ontology developed by the BRII project (O’Steen, 2008). Ontologies and taxonomies developed as part of the BRII project (and other endeavours at Oxford) are being stored and made available for others to use at http://www.vocab.ox.ac.uk. Controlled vocabularies and thesauri being used assist with research categorisation, search and retrieval. They include the Library of Congress Subject Headings (LCSH) as linked data (LoC, 2009) and the Humanities and Social Science Electronic Thesaurus (HASSET) thesaurus (UKDA, 2009). This will support the work of strategists and disseminators. Software is being created as part of the project that enables the easy construction of new vocabularies. This means that researchers with specialist expertise can create a vocabulary automatically applying the semantic web RDF structure, custom built for their research area which can then be used by others. This feature may be useful for strategists who want to label research entities in a way that is useful to them. Every entity is assigned temporal data. This tags when data was valid and when it changed and thereby describes the provenance of that data. The temporal data being recorded currently comprises three descriptors: ‘validFrom’, ‘validUntil’ and ‘validAt’. This helps demonstrate the trustworthiness of the data. In addition, data changes over time, but the institution needs to be able to report on previous status such as a project that ended 10 years ago. Each term and entity is resolvable to a URI where it is defined to ensure the context and meaning of the term is clear. Entities and terms are given as linked data to encourage further linking and innovative re-use. Use of this model is not yet fully understood by strategists and disseminators who are familiar with standard databases. It is likely to be some time before this group accepts the benefits of this method of data publication. 5. Benefits beyond Oxford Although the BRII project work is based wholly at the University of Oxford, there are a number of developments and services that will be made available for the benefit of the sector. Aset of software will be made freely available to enable others to create a vocabulary store similar to the http://Oxfordvocab.ox.ac.uk store. The vocabularies held in this store are available for others to use under a Creative Commons licence. The software described above in Section 4 enabling easy creation of vocabularies will also be made freely available. The collection of new and adapted vocabularies stored at http://vocab.ox.ac.uk will be freely available for use. Oxford is collaborating with other groups involved in similar and related work which are likely to result in tools and other developments that benefit the sector. Sharing expertise and experiences is feeding into national developments as these groups participate in discussions about research information management as part of a series of meetings organised by JISC (2009). The national focus of this work for research management information sharing is of interest to strategists and disseminators, particularly in connection with REF reporting. 6. Conclusions The BRII project provides an innovative solution to a problem that affects users across the Collegiate University in addition to the benefits for the institutional repository. It has been fortunate in its timeliness both in terms of concurrency with technical system and research management developments at Oxford and research information management activities taking place at the national level. There have been a number of influential internal ‘champions’ who understand how the registry has the potential to support and improve the working practice of strategists and disseminators across the university and assist the publicity of its research for the benefit of the whole institution, as well as individual academic divisions and central units. The critical success factors for the BRII registry (or probably more immediately, the Oxford Blue Pages) are ease of use, comprehensiveness, accuracy and keeping data up to date by way of regular data harvests from contributors. The project team needs to demonstrate the future likelihood of success in these areas in order to gain continuing support for development. The iterative mode of development and comprehensive user testing assisted user buy-in and the development of a user-friendly web interface. Use of the semantic web has been a vital factor in the development of the registry and its usefulness to strategists and disseminators. Each individual group has its own drivers and critical factors and it is this ability to reflect both individual preferences and the resulting complexity without trying to impose order that is decisive.

Health professions education (HPE) has entered a new era. Anonymous lectures in darkened halls and unstructured apprenticeships of varying quality are disappearing. Training future health care workers is evolving into a highly structured process with carefully formatted objectives, interactive teaching strategies, and multimodal evaluations.1,2 Institutions and accrediting bodies have established thresholds of achievement and proficiency as milestones for graduating learners to independent practice. 3,4 The number of questions arising regarding the quality and efficacy of educational practice has fed a growing field ofHPE research. This growth in HPE and a desire on the part of nursing and medical educators to disseminate their work have raised equally important questions about the ethical conduct of such work, particularly with the complexity and conflict presented for subject-learners, external funders, and regulatory demands. At the center of the debate is the institutional review board (IRB) and its proper role in the oversight of HPE. Background Traditionally, most educational program design and evaluation have not been considered part of the research enterprise. Clinicians or researchers would spend time teaching, but this was considered wholly separate from their other academic pursuits. The work of Glassick5 in the late 1990s provided a taxonomy and framework for a more deliberative teaching process with key components of careful design, meaningful evaluation, peer review, and reflective critique intended to spawn opportunities for presentation and publication. With the evolution of the ‘‘scholarship of teaching,’’ educators began to reflect on the research-like attributes of their work as well as the implications of its design and delivery for their subjects. In 2004, an article appeared detailing the plight of a grant-funded curricular programin geriatrics.6 The article described how medical students at a university in which faculty initiated curricular change with grant funding ultimately were forced to cancel the program when students protested the fact that they were not consented prior to the innovation and that the project had not undergone IRB review. While lacking the historical weight of the Belmont Report, this cautionary tale raised a number of important points, including (1) learners may be considered not just passive participants but actually subjects of education research and that educators need to consider their rights and identities when participating in these projects, (2) a lack of training and awareness among health professions educators about the protection of human subjects and the IRB process and a prevailing sense that IRB review is an optional process and something to be avoided if possible, and (3) the growth of external funding raising the possibility that local activities might be influenced by demands and requirements of outside agencies. These points speak to the core of the mission of the IRB in that health professions educators need to consider the ethical implications of the design and execution of their work. Significance HPE is recognized as a key ingredient in high-quality nursing andmedical care.7 A demand exists for evidence to inform how we educate health care professionals, and educational programs have become a key tool for implementing improvement. Admittedly, the science underlying education research is often complex and requires use of a variety of study designs—both experimental and quasi-experimental—andmultiple approaches for evaluation, including both qualitative and quantitative.8 The evolving sophistication and complexity of HPE research point to a need for a centralized source for review of conduct and quality of these projects.Moreover, education researchers are gathering, collating, and analyzing large amounts of data. These data need to be collected and stored in a way that ensures its integrity and protects the identity of learners. Collaboration across institutions adds an additional layer of complexity to this process.9 It is often the intent of medical and nursing education researchers to disseminate their work through presentation and publication. This intent to generalize has important implications for the use of information and protection of the identity of participants. Furthermore, HPE is moving toward measuring impact of teaching interventions on outcomes including patient care. The examination of a learner’s performance in real clinical care requires a consideration of howto handle not only the learner data, but also that of the patient aswell. Finally, consideration must be given to the learner as a vulnerable subject, prone to coercion, exposure, and undue psychological stressors. The IRB Process Health professions educators may be best served to reframe these challenges as compelling reasons to engage the IRB as a partner in the process of ensuring sound design and the responsible and ethical conduct of educational research. This ‘‘reframing’’ starts with a careful examination of the principles underlying the formation and mission of the IRB in the Code of Federal Regulations and a point-by-point examination of their relevance to the education researcher. Types of IRB Reviews The responsibility of an IRB is to ensure that the regulatory criteria for approval of a study are met and in so doing help protect the rights and welfare of human subjects involved in research, including ensuring that risks of participating in a study are reasonable in relation to the potential benefits. Current federal regulations require that all nonexempt human subjects research be approved by an IRB before an investigator may conduct a study. Making decisions regarding the IRB review process for research studies begin with first determining if human subjects are involved and if the investigation is truly designed to contribute to generalizable knowledge.10 The Federal Policy for the Protection of Human Subjects Research is known as the ‘‘Common Rule’’ and guides oversight and provides guidance for those engaged in research involving living individuals. 11 Human subjects research means obtaining data through intervention or interaction with an individual or by using identifiable private information to answer a research question.10 In the case of HPE, learners may be considered ‘‘subjects’’ when considering their rights and protections as participants. There are 3 types of IRB reviews for new protocols: exempt, expedited, and full-committee review. A full description of these reviews is found in Table 1. The IRB Review Process for Education Research There are specific categories of research involving human subjects that are eligible for exemption; some types of education research meet the definition for these categories. To assist education project leaders in determining what types of projects meet the exempt education research criteria, the Office for Human Research Protections in section 45 CFR 46.101(b1-6)10 includes educational research categories that could potentially be considered exempt: & Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods (CFR 46.101.b1). In Table 2, example 1 is an exempt education research study. & Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview, or observation of public behavior, unless (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects’ responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects’ financial standing, employability, or reputation (CFR 46.101.b2). This exemption does not apply when surveys, interviews, or observation of public behaviors include individuals younger than 18 years when the investigator is participating in the activities being observed. Research involving the collection of existing data, documents, or records, if publicly available, or if information is recorded by the investigator in such a manner that subjects are not identifiable (directly or through identifiers) may also be exempt (CFR 46.101.b4).10 Careful evaluation of the above categories determines whether an education research project meets the exemption criteria or requires an expedited or fullcommittee IRB review. A complete description of potentially exempt research categories can be found at http://www.hhs. gov/ohrp/humansubjects/guidance/45cfr46.html. When a research study does not meet the criteria for exemption, the risks to subjects participating in the project should be further evaluated. In Table 2, example 2 is of a nonexempt education research study. Federal regulations define minimal risk as the ‘‘probability and magnitude of harm or discomfort anticipated in the research is not greater in and of themselves than that ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests’’ (45 CFR 46.102[i]).10 Most education research studies are minimal risk studies when protection of the rights and welfare of subjects involved are carefully considered. Education research subjects must not be unduly influenced to participate.13 Protection of the identity of involved subjects must be considered. If any information is collected that can directly identify a subject or when an identifier can be linked to this individual, then it is not eligible for exemption unless it meets one of the 6 exemption categories listed in the federal regulations (45 CFR 46.101).14 Measures to protect the privacy and confidentiality of subjects should be taken and clearly described in the study. When students are directly involved themselves, even exempt education research studies must inform the subjects about the project, and they must be provided details as to how confidentiality of the information will be protected. Subjects must also be told that participation is voluntary and that they can refuse to answer questions that they do not wish to answer. Subjects should be allowed to withdraw at any time without repercussion. Written consent may be required if the project includes collection of recorded data via voice, video, digital, or image. Individual or group interviews or focus groups may require written consent as well. First consider whether the research is more than minimal risk of harm to subjects and if it involves no procedures for which written consent is normally required outside the research environment (45 CFR 46.117.c2). There are many situations where consent documentation can be waived, and the IRB can request the investigators to provide subjects with a written statement regarding the research. Several examples are provided to give further guidance. Examination of nursing students’ perceptions about a specific disease such as human immunodeficiency virus (HIV) or cancer and their attitudes toward caring for a patient through interviews or focus groups requires research measures to ensure that student participation is voluntary and anonymity is protected. This type of research would not be exempt and requires informed consent. In comparison, evaluation of the effectiveness of a curriculum designed to allow students to explore their attitudes toward caring for patients with HIV or cancer may be seen as standard educational practice and would meet the criteria for exempt research if anonymity is maintained and no identifiers are collected. Even if written consent is not required, learners responding to surveys or other measures should be informed of the voluntary nature of their participation and the intent of the educator to analyze and disseminate the data. Improving the Process HPE researchers and their supporting institutions aspire to innovate in a way that will improve both the ability of the graduating learner and the care they provide. To achieve these goals in a responsible and ethical way, steps must be taken to bridge the gap between their work and the review process. These improvements may take several forms including (1) education and orientation programs, (2) expansion of the types of expertise and assistance available at the IRB, and (3) development of HPE-friendly review templates and processes. Institutions should begin by introducing programs to inform both educators and IRBmembers about issues specific to education-related projects. Educators need to gain greater understanding of the process and definitions set forth in the CFR to more accurately prepare submissions for review. Specifically, such programs might focus on criteria for exemption versus review, risk for learners, and indications for consent. Health professions educators would also benefit from more in-depth orientation to data management and protection and, specifically, institutional resources designed to help with these issues. Finally, clear guides to the logistical aspects of submissions are essential. The IRB itself also should participate in a review of its capabilities for reviewing educational projects. This may involve orientation for its board members on the specific issues and characteristics of education research. In addition, many have proposed that IRBs should seek to include reviewers with specific expertise in social and behavioral sciences (often lacking on medical campuses).15 At institutions with active education research communities, addition of a dedicated chair or board may be necessary. In addition to orientation and education, the development of submission templates and targeted reviews specific for education projects may help the utility and efficiency of the process.16 At Duke University, a group of educators from our Schools of Medicine and Nursing developed a template in collaboration with leadership of the IRB to address needed improvements in the process. Our work followed the example set by the development and implementation of a template for submissions for quality improvement (QI) projects.17 Our team first met with representative chairs of the IRB to discuss the structure and content of the template as well as a plan for piloting the submission and review process. A draft version of the template was circulated among educators and IRB members for feedback and revision. The template, derived from the traditional research submission form, offered a modified organization and language to provide a better fit for education-related projects as well as clarity for reviewers (Table 3). A pilot group of submissions was routed to 2 IRB chairs involved in the development of the template. The template proved quite popular among educators and reviewers in both form and function. The form was posted on the IRBWeb site, and our development team offered a series of workshops for faculty on its use. Conclusion It is important for health professions educators to consider the ethical implications of conducting education research. Increased understanding of the categories for exempt research and how to determine when an education research study does not meet the criteria for exemption are essential for those involved in HPE. Specifically, educators need to recognize learners as subjects and anticipate and acknowledge risks inherent in their participation in education research. In addition, educators need to seek greater knowledge and resources for data collection and management to ensure protection of subject identity and study integrity. Improvements in the review process for educational research studies will require a collaboration between educators and the IRB as outlined in this article. Development of education programs for researchers and reviewers, integration of educational expertise on the IRB, and standardization and dissemination of submission templates will all enhance the process and improve the quality of the work. References 1. Cooke M, Irby DM, O’Brien BC. Educating Physicians: A Call for Reform of Medical School and Residency. San Francisco, CA: Jossey-Bass; 2010. 2. Broome ME. Building the science for nursing education: vision or improbable dream. Nurs Outlook. 2009;57(4):177<179. 3. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. N Engl J Med. 2012; 366(11):1051<1056. 4. Pijl-Zieber EM, Barton S, Konkin J, Awosoga O, Caine V. Competence and competency-based nursing education: finding our way through the issues. Nurse Educ Today. 2014;34(5):676<678. 5. Glassick CE. Boyer’s expanded definitions of scholarship, the standards for assessing scholarship, and the elusiveness of the scholarship of teaching. Acad Med. 2000;75(9):877<880. 6. Tomoko wiak JM, Gunderson AJ. To IRB or not to IRB? Acad Med. 2004;79(7):628<632. 7. Institute of Medicine (US) Committee on the Health Professions Education Summit; Greiner AC, Knebel E, eds. Health Professions Education: ABridge toQuality.Washington, DC:NationalAcademies Press (US); 2003. 8. Broome ME, Ironside PM, McNelis AM. Research in nursing education: state of the science. J Nurs Educ. 2012;51(9):521<524. 9. Oermann MH, Hallmark BF, Haus C, Kardong-Edgren SE, McColgan JK, Rogers N. Conducting multisite research studies in nursing education: brief practice of CPR skills as an exemplar. J Nurs Educ. 2012;51(1):23-28. 10. Code of Federal Regulations, Title 45, Public Welfare, Department of Health and Human Services, National Institutes of Health, Office for Protection From Research Risks, part 46, Protection of Human Subjects. Effective January 15, 2009. Available at www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm. Accessed May 9, 2015. 11. Federal Policy for the Protection of Human Subjects (‘Common Rule’). Available at http://www.hhs.gov/ohrp/humansubjects/ commonrule/index.html. Accessed May 9, 2015. 12. Schoening AM, Selde MS, Goodman JT, et al. Implementing collaborative learning in prelicensure nursing curricula: student perceptions and learning outcomes. Nurse Educ. 2015;40(4): 183<188. 13. Miser WF. Educational research—to IRB, or not to IRB? Fam Med. 2005;37(3):168<173. 14. Sullivan GM. Education research and human subject protection: crossing the IRB quagmire. J Grad Med Educ. 2011;3(1):1<4. 15. Dyrbye LN, Thomas MR, Mechaber AJ, et al. Medical education research and IRB review: an analysis and comparison of the IRB review process at six institutions. Acad Med. 2007;82(7):654<660. 16. JohanssonAC,Durning SJ, Gruppen LD, OlsonME, Schwartzstein RM, Higgins PA. Perspective: medical education research and the institutional review board: reexamining the process. Acad Med. 2011;86(7):809<817. 17. Demeo SD, Nagler A, Heflin MT. Development of a health professions education research specific institutional review board template. Acad Med. 2015. In press.

Throughout the academic and research community, interdisciplinary research has become a catch phrase (Giacomini, 2004; Robertson, Martin, & Singer, 2003). With the recent emphasis in the NIH Roadmap initiative (http://www.ncrr.nih.gov/ roadmapnewsecir.asp) on interdisciplinary and translational sciences, interdisciplinarity has become the model of scholarly inquiry generally espoused by many who seek and receive federal research funding. Despite this, there are major gaps in our general understanding of interdisciplinary research and how it can be successfully integrated and sustained in academic health science centers and universities (Mallon & Bunton, 2005). Entities designated as interdisciplinary research centers abound in large universities and academic health centers, but in many settings the mantra of interdisciplinary research may be no more than lip service. Such centers have been described as follows (Committee on Facilitating Interdisciplinary Research, 2004): Some are bigger and intellectually more influential than some academic departments. Others are highly specialized and narrow. Some have existed for decades, others disappear after only a few years, and still others merge to create new units or emerge when one interdisciplinary unit is split. Some have retained their original purpose throughout their lifetimes; others have substantially shifted their academic focus. (p. 20) Considerable ongoing resources and efforts are being expended in these research centers. Although they are highly variable in their goals, administrative structure, funding, and defined outcomes, it is likely that there are also many commonalities and potential interfaces or even overlaps among them. Unfortunately, however, those characteristics that are predictive of success of such centers have not been clearly articulated or codified. Research centers are different from other academic units, and are relatively independent of the existing structure of a university. This means that they can undertake innovative research agendas free of the regulations of accrediting organizations, the routine activities inherent in administering educational programs, and the obligations of participation in university administrative activities. They are – or are intended to be – interdisciplinary, so that they can support research teams that cross disciplinary and departmental lines and their members can conduct research that falls outside the established bounds of a disciplinary department. Finally, centers are problem-responsive. They arise to confront specific issues and concerns, drawing together faculty whose work addresses these problems. Interdisciplinarity, independence, and responsiveness are the principle strengths and rationales for the existence of research centers. At the same time, these features present centers, and the universities that house them, with several distinct challenges. In this paper, we report on the results of a conference of directors of diverse research centers at a single research university that focused on the challenges facing centers and their universities and the factors predicting their success. The Conference In 2004, the National Institutes of Health allocated funds for exploratory centers in interdisciplinary research (http:// www.ncrr.nih.gov/roadmapnewsecir. asp). One of the 21 centers funded was the Center for Interdisciplinary Research on Antimicrobial Resistance (CIRAR, http://www.cumc.columbia. edu/dept/nursing/CIRAR/). CIRAR’s core research collaborative team includes persons from the disciplines of epidemiology, microbiology, pediatrics, infectious disease, nursing, economics, health policy, education, statistics, economics, informatics, and public health. The goals of this Center were not only to develop a research agenda that would have an impact on the global problem of antimicrobial resistance, but also to establish a vital, sustainable interdisciplinary research process. Despite the recognized need for interdisciplinary collaboration in biomedical research, there are structural and cultural disincentives within the academic setting that must be overcome. Hence, we developed a series of strategic initiatives to systematically examine the structure, processes, and outcomes necessary for an interdisciplinary research center to thrive. One of our first orders of business was to review bodies of literature from business, education and health care to adapt and develop our own definition of interdisciplinarity which could then be used to identify the competencies needed for successful interdisciplinary research practice. From this literature review an initial definition was developed and small modifications were made after field testing. We defined interdisciplinary research as any study or group of studies undertaken by scholars from two or more distinct scientific disciplines. The research is based upon a conceptual model that links or integrates theoretical frameworks from those disciplines, uses study design and methodology that is not limited to any one field, and requires the use of perspectives and skills of the involved disciplines throughout multiple phases of the research process. The process we used to address the definitional aspects of interdisciplinary research has been described elsewhere (Aboelela et al., 2007). Our second strategic initiative was to convene a group of directors of interdisciplinary research centers in a halfday symposium to accomplish five aims: (a) identify characteristics essential to successful interdisciplinary research centers; (b) assess challenges in the operation of a research center and strategies to deal with these challenges; (c) discuss mechanisms for sustainability of centers (e.g. funding); (d) increase networking and communication among interdisciplinary research centers; and (e) exchange successful strategies for enhancing minority and gender balance in interdisciplinary research centers, as well as the balance of junior and senior researchers. Because no list of such centers existed at the University, we searched websites and polled departments and schools to identify relevant centers, using the following criteria: the center had to be interdisciplinary with a major research mission and have current external funding from the government, foundations, and/or professional organizations. We identified 65 centers across Columbia University that met these criteria and contacted directors either directly by telephone or email. While there was some initial skepticism among directors and academic administrators about whether such a meeting would yield a useful outcome, the majority of center directors were enthused and supportive, noting that there was little opportunity for such interface. The forum was convened in November 2005 with 59 attendees from 29 different centers. Also in attendance was a project officer from NIH, the vice president of the university, and several deans. Eight center directors and two moderators, who also serve as center directors, formed two panels to lead discussions responsive to each of the aims of the forum, and there was considerable input from the entire audience. Three professional staff members took extensive notes, panels were audiotaped, and consistent themes were summarized at the end of the day by a skilled facilitator. Summarized below are the thematic challenges identified by participants, discussion regarding the interface of the centers and the university, and a summary of issues and recommendations that emerged from this conference. The Challenges Identified The following represents a qualitative summary of the discussion that ensued. Our review of the conference proceedings suggests that the challenges to success facing research centers fall into 3 categories (Table 1): fiscal sustainability, recruiting and retaining faculty, and leadership sustainability. Fiscal Sustainability Many, but not all centers at the university began with a substantial research grant. A small number began with funding from school or university administration or from an outside gift. This initial funding allowed the centers to become established and to embark on their programs of research, and also financed or enabled a request for space and other resources, such as administrative support. Over time, center financing evolved. Successful centers generally obtained additional outside grant support to continue or enlarge their research programs. These new grants, however, often raised challenges for the centers, especially when they were written by faculty from disciplinary departments who had joined the center. The new grants brought indirect cost recovery (ICR) funds, the distribution of which among the university, schools, departments, and the center itself had not always been clearly contemplated at the establishment of the center. Centers often required new resources – space, faculty, or administrative support – and center directors complained that obtaining these resources sometimes necessitated extensive negotiation. Policies with respect to the distribution of ICR funds varied considerably across the University. Center directors noted that the ability to maintain control of some ICR funds facilitated the task of maintaining the center over time. Centers with wellestablished protocols for sharing ICR with disciplinary departments also found that this practice brought them needed support from the departments. Centers without access to ICR funds, especially those without an outside endowment, had to develop strategies that would allow them to make longer term commitments to participating faculty. In some cases, centers experienced an interval between grants when funding was insufficient to maintain core resources. Generally, centers did not have guaranteed sources of bridge funds for these circumstances. Those larger centers that both held many grants simultaneously and obtained a share of ICR funds sometimes had some wiggle room, but centers with fewer grants found it difficult to set aside a share of funds (from whatever source) and had to negotiate bridge funding. Center directors agreed that reliance on direct federal grant funding alone was problematic. They noted that having a diversified portfolio of financial supporters (including a combination of government, industry, foundations, endowment, and university funds directly or through ICR) helped provide stability. Recruiting and Retaining Faculty The initial development of a center generally required identifying faculty across disciplines with an interest in a topic area. Successful centers had identified research areas where there was a widely shared sense of need for more collaborative work. Several center directors remarked that they had been flooded with requests to participate when the center was first developed. Challenges around faculty arose for three reasons. First, the center directors agreed, excellent disciplinary researchers committed to a problem area and excited by the prospect of collaborating with others may nonetheless fail to thrive in an interdisciplinary research environment. Centers depend on faculty who are both rigorous scholars and can function well in an institutionally unusual environment. They must be willing to learn the language and constructs of other disciplines. They must have, as one center director put it, a high level of intellectual curiosity, tolerance for ambiguity, and ability to play with others. Center directors struggled with identifying such individuals and with the problems created by members who did not fit this bill. Some faculty members were simply not interested in spending the time necessary to work across disciplines or sharing their perspectives and research interests with others, i.e. they were not cut out for an interdisciplinary environment. Many found that younger faculty members were more malleable and fit into the center better than did more established scholars. The need to satisfy disciplinary department promotion criteria, however, can make participation in an interdisciplinary center difficult for junior faculty. Moreover, centers cannot function exclusively with young faculty. They need more senior faculty members to act as “heavy hitters” and obtain substantial grants, as well as to manage the administrative tasks of the center even though some may be less accommodating than junior faculty. Second, centers needed to retain and replenish the ranks of their faculty over time. Center directors needed strategies for faculty recruitment and retention throughout the life of a center. They reported that the establishment of core facilities often acted as a magnet that drew and held faculty to the center. Many centers offered pilot grants and seed money to investigators. Moving beyond pilot projects required new kinds of collaboration and communication among center members. Conference participants pointed out that such communication can be difficult. For example, the culture of the private sector where interdisciplinary collaboration has been most successful emphasizes discovery and application of profitable products, while academics may be more interested in mechanisms and new discoveries. In other cases, collaborators may have very different styles of communication, as well as different perspectives on sharing and ownership. Because of the nature of the work, some disciplines may have varying vocabularies and methods, expectations about the pace or hours to be worked and standards of proof. Some investigators favor rapid publication of each new finding; others prefer to amass a body of work for a single large publication. Some are open to large teams and data sharing while others prefer to minimize interactions. Thus, working and communication styles played important roles in attracting or failing to attract and retain faculty over time. The need to recruit new faculty often generated a third problem. At this university, as at most others, only disciplinary departments may make faculty appointments and promotions. In some cases, centers may appoint researchers using non-professorial titles. Several center directors noted that these titles were less valued in the university than traditional titles. Center directors often needed to work with disciplinary department leadership to recruit faculty who were expected to participate exclusively in center activities. One center director suggested that permitting joint appointments between a department and a center might facilitate such recruitments. In some cases, centers draw in most of the faculty of a given disciplinary department. The center may saturate a department with faculty. In these situations, the boundaries between the department and the center may disappear altogether. One university administrator noted that in this situation it might make more sense to convert the center into a department of its own. Leadership Sustainability The final set of challenges facing centers concerned leadership. Center directors must be charismatic advocates for their research areas and for the enterprise of interdisciplinary research. They must be able negotiators, finessing arrangements with university administrators, department chairs, and both accommodating and less accommodating center members. The nature of interdisciplinary work means that they must do all this in a collaborative rather than a dictatorial style. Finally, they must be skilled administrators. Several directors understandably complained that the administrative demands of managing a center were very time consuming. Centers are generally developed because an individual with this rare combination of qualities initiates them. Problems may arise over time, however, when these pioneering leaders seek to share the burden of management or leave their positions. Center directors noted that new leadership was likely to be drawn from the ranks of senior center members who viewed this role as a professional obligation. Centers and the Institution All three of the challenges we identified arise from the problem of establishing the natural lifecycle of a center. Problem responsive centers are fundamentally different from existing university institutions. They occupy a place between academic departments and individual grant-funded projects, both institutional forms with well-understood lifecycles. Our university, and we suspect most others, does not have established criteria for defining when centers should be established, how they should be sustained, and when they should be closed. Individual grants are initiated by faculty and usually managed in the context of an academic department. They begin on the funding date and end (usually) when the grant expires. Financing, personnel, and leadership throughout the grant period are clearly specified in the grant proposal and funding statements. Departments are developed very slowly. Generally, the formation of a department requires several layers of academic approval from the school, the university administration, the faculty senate, the board of trustees, and sometimes the State. To initiate a department, a school must clearly define the discipline represented, the teaching need and academic mission, and availability of appropriate resources to meet the articulated needs. Once established, a department is built on the financial and scholarly bedrock of its teaching mission. Sufficient faculty must, at the very least, be retained to teach courses required by accrediting agencies. These agencies, in turn, provide an outside force prompting the university to maintain the viability of the department. Teaching revenue, while often limited, provides a stable backstop against volatile outside “soft money” funding. Closing a department, a very rare event, likewise requires a series of steps, and the academic institution usually remains responsible for compensating any tenured faculty in a department that is closed. Demands from students, accrediting agencies, and others, and the existence of teaching revenues, require that universities have well-established procedures for evaluating and maintaining their academic departments. Procedures exist to recruit faculty when positions become available, and to promote faculty through promotions committees. Universities also have procedures for recruiting departmental leadership, whether through a system of rotation or a search process. Finally, most universities have formal systems of departmental review, during which outside committees periodically assess the performance of each department. Centers fall somewhere between individual grants and departments. They begin with much more university buy-in than would an individual faculty member’s grant proposal. Since interdisciplinary research centers exist to address a new area of research, they do not require all the steps needed to establish a department. Centers generally have a specific mission statement and aims defining the proposed scope of the center. Unlike the case of a grant, however, this statement generally does not specify when the work of the center will be completed or what the criteria would be to close the center. Research centers, unlike academic departments, often do not collect teaching revenue. Most depend on the school or university administration to help them maintain fiscal sustainability, either through ICR sharing or direct commitments. Without pre-specified guidelines about what constitutes center success and what the university’s commitment to the center will be, center directors cannot always rely upon these potential funds. This lack of dependable funding leads center directors to seek independent endowment support. This, in turn, can pose challenges to the university if the rationale for the existence of the center no longer exists or if centers compete with other university priorities for outside funding. Further, centers usually do not have a natural constituency, unlike departments, which can depend on their current students and alumni, as well as accrediting agencies, to advocate on their behalf. Several of the center directors at our conference spoke of their efforts to develop a constituency in the outside community to provide them with leverage as they built their centers. An outside constituency relieves some of a center director’s burden to continuously justify the university’s commitment. There is generally no established procedure for sustaining leadership in centers. In the case of a single grant, the life of the grant is coincident with the participation of the lead investigator. In the case of a department, the existence of the department is independent of the present leadership. In the case of a center, leadership and existence are intertwined. If a university has no systematic procedures for deciding when a center is successful or should be perpetuated, the decision to maintain leadership for a center is made separately in each case. Centers should not be departments. They should come into and out of existence more easily and fulfill missions that departments cannot. But as centers become an increasingly important component of the university’s institutional life, more formal procedures are needed to monitor their establishment, continuation, and termination. These procedures will help the university control its overall operations and ensure the quality of the centers. It will also help center directors, who will be able to rely on a set of defined privileges and obligations as they strive to build their faculties and research programs. Summary and Recommendations The process of collaboration requires institutional and individual commitment, but formal partnerships such as research centers are regulated primarily at the institutional level. Nearly all institutions have rules and guidelines for interdisciplinary research to govern ownership of work products and data, material transfer, and academicindustrial agreements. In general, external collaboration cannot proceed without involving the institution. Although guidelines or regulations do not explicitly cover many aspects of collaboration, the goal should be communication that clarifies expectations of all parties involved. For these reasons, policies, procedures and principles for management of interdisciplinary research centers need to be explicit. The challenges of interdisciplinary research centers highlighted by participants in this conference—fiscal sustainability, recruiting and retaining faculty, and leadership—have been recently summarized in a report published by the National Academy of Sciences (2004). To our knowledge, however, our symposium was the first formal meeting of a large cadre of research center directors to address the aims we articulated. While there remain at many universities structural challenges to interdisciplinary research (e.g. policies and processes for sharing of ICR funds), we recognized that the major challenges as well as the major sources of gratification associated with research centers are interpersonal as well as institutional. This conference served to facilitate and support an institutional shift towards an environment in which interdisciplinary efforts thrive. This is well within the ethos of the university whose faculty strive to work in collaboration with those outside of their own disciplines. Following this conference, a senior staff member was hired by the university to focus on the development and support of interdisciplinary research. Based on the proceedings of the conference, we make the following recommendations for institutions in which interdisciplinary research centers are housed:(a) maintain a database of interdisciplinary research centers within a centralized office (e.g. grants and contracts or research office) for the purposes of networking and tracking; (b) provide an ongoing forum for interaction among directors and members of interdisciplinary research centers; (c) establish criteria for defining when centers should be established, how they should be sustained, and when they should be closed (i.e., what the natural lifecycle of a center should look like); (d) clearly identify individuals/offices within the institution that are responsible for policies regarding issues such as indirect cost sharing, faculty recruitment into centers and/or departments, and other administrative policies that influence center operations and success; (e) provide support for development of interdisciplinary leadership skills; (f) develop formalized mechanisms to assure that interdisciplinary activities are acknowledged and rewarded in the faculty promotion and tenure process; and (g) explore the role of interdisciplinary centers in developing and contributing to coursework designed to prepare researchers with interdisciplinary expertise.

Institutional research is a term that is used in reference to a wide variety of activities ranging from data processing to disciplined inquiry. Several people have attempted to define institutional research--Maasen (1986) in terrns of the main data handling activities; Peterson (1985) in terms of its linkages between institutions of higher education and the functions of higher education; Saupe (1981) in terms of its use in planning, policy analysis, and decision making. Definitions have failed to characterize important aspects of institutional research. "This . . . has become something of an embarrassment to those of us who have spent a number of years calling ourselves institutional researchers" (Muffo and McLaughlin, 1987, p. iv). The attempts at definition fall to address the process and structure of the research part of institutional research. An "ideal type" of disciplined inquiry has yet to be identified to guide the practitioner in doing institutional research. "Ideal type" is a term used by Ma× Weber (1949) to describe an abstract utopian construct that characterizes a set of relationships. "It is not a description of reality but it aims to give unambiguous means of expression to such a description" (Weber, 1949, p. 90). A logical structure or model would be an ideal type. Like a map, it communicates structural essentials, thus reducing ambiguity. Having an ideal type allows one to better know how to go about doing what needs to be done. When one speaks of disciplined inquiry or research, the natural assumption is that one is talking about science. Science provides the ideal type that comes to mind when such things are discussed. For example, Von Vught (1990) compared institutional research to science, and argued that institutional research faces legitimacy problems with the scientists that make up the academic community. (Inherent in this argument is that institutional research should be compared to science, and in fact should be science.) Upon reflection, we find at least three ideal types associated with science. The first is the structure of the process of modern science, usually referred to as scientific method. The second is the logical structure of science. This ideal type deals with scientific ways of thinking. The third ideal type is related to the social structure of science and the social elements that contribute to the growth and development of an understanding of a set of related phenomena. This paper outlines the ideal types mentioned above and discusses institutional research's characteristics in relation to these structures of science. The first section of the paper examines scientific method as a process of disciplined inquiry, then compares institutional research to that process. The second section compares the logical structure of institutional research with the logical structure of scientific inquiry. The third section utilizes Thomas Kuhn's (1962) notion of scientific paradigms to examine the community of institutional researchers. It is hoped that this discussion will lead to identification of ideal types in institutional research, dispelling misconceptions about the "research" in institutional research and resulting in improvements in the quality of work done in the area. SCIENTIFIC METHOD Like institutional research, science as a form has itself defied simple definition. Philosophers of science have for centuries refined and reshaped the concept of science, and have argued about its features as a disciplined method of inquiry. At different times science has meant different things. It has undergone paradigm shifts--dramatic alterations of focus and perspective. With each shift, it has incorporated earlier versions of science to some degree, but has taken a new approach in order to correct for weaknesses perceived in the previous paradigm. With each new approach, either the previous paradigms have been abandoned or they have been accepted as tools to be used in the quest for knowledge. In spite of the dynamic natures of science and institutional research, or perhaps because of them, science and institutional research have a great deal in common. SCIENCE & INSTITUTIONAL RESEARCH "Scientific method" is a logical process with which institutional research shares several features. Scientific method consists of the following: observations, facts, hypotheses, experiments, theories, and laws. The logical structure begins not with theories, but rather first with observations and facts (Casti, 1989). Observations are more than just catalogued events; they often include groupings of data. Observations are made that single pieces of data group in a way that suggests to the observer that the grouping is meaningful. Conjecture about the explanations for the patterns observed leads to hypotheses. Experiments are conducted to test the hypotheses. If a hypothesis is supported by the evidence, the identification of laws and theory building follow. Institutional research follows a similar logical process. Observations and facts serve as the foundation for disciplined inquiry. Data and facts about the institution and about higher education are examined for patterns. It is the observers (which may include administrators and faculty, as well as the institution at researcher) who decide which information is relevant to the question at hand. The way in which the question is posed influences the understanding of what is relevant. The environmental factors that influence decision making also determine what is relevant. Pattern identification, the result of repeated observations in a context created by fact, is an important part of institutional research. (Sometimes the institutional research function ends here and others interpret the information as they see fit, but not always.) As with science, conjecture about the patterns of data can lead to hypotheses. "Experiments" are sometimes conducted to test the hypotheses. However, that is a point at which science and institutional research diverge, at least in the eyes of the traditional scientist. Laws and theories have not been identified for a discipline of institutional research. This may be due to the fact that few attempts have been made to look for common principles that underlie inquiries into the same types of phenomena. However, some argue that there are no true laws associated with human social behavior (see Nagel, 1961, pp. 447-546). A law is "a statement of a relation or sequence of phenomena invariable under the same conditions." Theory is defined as "a group of general propositions used as principles of explanation for a class of phenomena" (Random House College Dictionary). lnstitutional research does not at this point have laws and a body of theory that are specific to a particular discipline. However, institutional research does share some assumptions with the social sciences. They are as follows: 1. Human behavior can to some extent be observed, measured, and predicted. 2. There are environmental factors that influence behavior. 3. Some environmental factors that influence behavior can be manipulated. 4. Manipulation of environmental factors, including social factors, can lead to changes in human behavior. 5. Certain environmental and social factors have more influence on human behavior than others. 6. Social factors associated with certain characteristics of individuals may influence the ways in which the individuals behave. 7. Many of an individual's characteristics associated with behavior cannot be changed. As in the social sciences, many of the relationships identified in institutional research are probabilistic in nature (see Nagel, 1961, pp. 22-23). "Though the premises are logically insufficient to secure the truth of the explicandum, they are said to make the latter 'probable'" (Nagel, 1961, p. 22). One reason laws and theories of institutional research have not been identified is that they have been looked for in the wrong places. Laws and theories are usualty in the domain of the empirical. However, the laws and theories of institutional research lie in the realm of the normative. Some of social science is involved in the process of envisioning an ideal, what could be, given the right circumstances. An example is democratic theory, the utopian ideal of government that consists of the following three elements: popular sovereignty, majority rule, and equality. Political scientists make observations about how and to what extent governments approximate the utopian ideal of democracy. Institutional researchers often operate under the assumption that there is an ideal condition, predetermined by the mission, goals, and policies of their college or university, or by the immediate needs of the administration. They attempt to find the appropriate information that will contribute to understanding and working toward that ideal. The policies, mission, and goals of the particular institution serve as the body of theory for that institution (not for a discipline of institutional research). We will catl this the theory of the institution. The functions of institutional research are geared to this level. The theory of the institution is concerned with the understanding of the institution's workings and identity. Strategic planning and management are the processes through which new policies (potential contributions to the theory of the institution) are tested for their ability to contribute to reaching the ideal as stated in the institution's mission and goals. They make up the normative theory of institutional research. We will call this the theory of institutions of higher education. This body of theory is concerned with the processes used to arrive at findings, rather than the findings themselves or a generalization of findings. Laws and bodies of theory have not been identified in institutional research partly because of the types of questions institutional research attempts to answer. Science in general is intended to respond to the question "why?" However, institutional research in its capacity of decision support rarely approaches an issue from an explanatory perspective, but rather responds to the question "which?" When an issue arises concerning a policy dec'ision, the decision makers want to know which choice is the best one when all of the facts available are taken into account. Included within the "which" questions are often smaller questions: "who? .... when? .... where? .... how much? .... did it work?" and "does it matter?" Science, on the other hand, attempts to discover a cohesive set of truths about the nature of reality. In science there is no intent to solve practical problems, only attempts to contribute to knowledge. A useful analogy may be to imagine that science is an investigation into how a tool functions or why it works. For example, mathematicians, in their research, may focus on how and why a particular type of statistical operation works. Institutional research, in contrast, assumes that the tool works and applies the tool for its intended purpose to produce something that will be used within a specific real world context. Science creates knowledge, while institutional research goes one step further by using the knowledge created through science to meet practical ends. Institutional research is science applied. Earlier, we mentioned that institutional research does not have laws and theories in the same sense that we think of science as having laws and theories. This is true, but there are some things that are invariable and act as principles for explaining sets of phenomena in institutional research. Underlying the mission and goals of any institution are two things that do not change in concept: intent toward student success and the maintenance of organizational viability. These figure into the decisions made in an institution of higher education. How student success is defined or measured rnay depend on the culture of the particular institution. The strategies used in attempts to maintain the viability of the organization are likely to differ from institution to institution and from one time to another. The constraints on the institution may be location or constituency specific. Nevertheless, all institutions of higher learning share the assumptions that encouraging student success and maintaining organizational viability are inviolable principles. In the place of theories in institutional research are the assumptions õf what is appropriate, important, and beneficial to the unending pursuit of student success and organizational viability. Sometimes, the ways in which these are characterized or envisioned change with what is in vogue in higher education, or with the socially or politically popular issue of the day. Planning to cope with faculty shortages, providing incentives for increasing cultural diversity, and buffering against budget turbulence are examples of such current issues in institutional research. Even though these driving principles of institutional research do not at this point make up a cohesive body of knowledge as do laws and theories in science, institutional research is not far removed from science. Any one of the topics of study of institutional research can be broken down into topics of narrow scope suitable for scientific study, and can be investigated using the laws and theories of one or more of the social sciences for the purpose of adding to understanding of how the world works. Some argue that they should be in order to improve the validity of institutional research. THE STRUCTURE OF SCIENCE The logical structure of science, as we know it today, is based on deduction or induction, and experimentation, modeling, and refutation. Each of the building blocks of the form of science was the result of an attempt to correct for flaws in previously used methods of explaining reality. Deduction was the "science" of Aristotle's day. In deduction, premises are postulated in such a way as to lead to a conclusion that is supported by the premises. The truth of the premises is simply assumed. In institutional research, the assumptions more than the facts are responsible for many conclusions. The truth of the premises that information kept on databases about students, faculty, facilities, and budgets provide meaningful and appropriate answers is assumed. That the mission and goals of the institution are appropriate to achieving student success and ensuring organizational viability is usually assumed. Most important, it is assumed that applying conventions of science to decision making leads to better decisions than reading tea leaves or flipping coins. Deduction moves from generalities to conclusions about a specific. The most famous example of a deductive. argument follows: All men are mortal. Socrates is a man. Therefore, Socrates is mortal. Although deduction is less used in institutional research than induction for arriving at conclusions about the conditions with which the institution must cope, it is not uncommon. As example used in institutional research might look something like the following: All students must be able to read and write well enough to pass a minimum skills test in order to succeed in school. John Doe cannot pass the minimum skills reading and writing test. Therefore, John Doe cannot succeed in school. Such assumptions may be a part of an assessment procedure used for admissions purposes or for evaluating the effectiveness of a program. The test's validity (of various types) is assumed. The relationship between test performance and success in school is assumed. The truth of the premises is based on a set of decisions made in an artificial environment instead of being based on assumed truths about the nature of reality (but philosophical arguments about differences between created reality and perceived reality are beyond the scope of this paper). Induction, a concept advocated by Francis Bacon for understanding the reality of the world through the facts rather than through preconceived ideas about the nature of reality, relies on repeated observations of facts. These observations are then used as the basis for predictions. The following is an example of induction: I have observed that for 30 days the sun has set in the west. Therelore, based on these observations, I predict that the sun will set in the west tomorrow. In institutional research, induction is essential to informing decision making. Trends are frequently used for projecting such things as enrollments, financial needs, market share, attrition, and other factors that colleges and universities must adjust or adjust to for the sake of maintaining viability. Galiteo brought a new perspective to science when he introduced the controlled experiment. Controlled experiments improved upon induction and deduction by narrowing the number of possibilities in establishing causation. With deduction, the relationship was built into the argument as an assumption. With induction it was assumed that a series of like events had the same cause, although causation might never have been addressed in the process of inquiry. The modern conception of science began with the introduction of study of causal relationships. Although institutional research makes use of experimentation to some degree, it is like the majority of the social sciences in that it is dependent on using quasi-experimental and nonexperimental designs for attempting to understand the relationships between events. More often than not the best that can be achieved in the social sciences is the establishment of a probabilistic relationship between a group and one who shares certain characteristics of that group (Nagel, 1961). Institutional research suffers from a variety of methodological problems that go beyond even those of the social sciences. Orte major difficulty with which institutional research taust deal is that the entire population is offen the sample upon which inferences are based. Another is that obtaining sufficient information about the possible influences on events is offen too costly or too time consuming to be of benefit to the institution. The greatest problem is the reactive nature of institutional research. It is often the school's administration that decides which variables would be most useful to understanding the relationships between observed events, possibly resulting in "looking where the tight is best," rather than looking where the answers are most likely to tie. The formulation of the question determines the kinds of relationsips examined. Modeling has added to science by formalizing a method for representing the relationships between observations. Modeling reduces problems of interpretation of relationships. It provides a "language" upon which all can agree, thereby making it possible for others to converse about relationships between observations in a way that is free from prejudice. Formal mathematical models and statistics and diagrams are the languages used to describe the nature of relationships. Modeling often provides the most convincing argument for choosing one policy alternative over another institutional research. Modeling, like scientific method, reduces the infornaation necessary for understanding to a manageable amount. Although the decision makers in the institution rnay have little understanding of statistics, they may understand a few key concepts, such as significance, correlation, and reduction of error. They may respond to charts and diagrams better than numbers. Once a model is built, it may be used repeatedly for evaluating and for projecting. It is the use of statistical techniques that leads some who participate in institutional research to believe that they are practicing science. Knowledge of and skill in the use of sophisticated techniques are erroneously equated with an understanding of how science works. However, statistics and formal mathematical modeling are simply tools. They can be used in the pursuit of scientific knowledge and in the process of arriving at practical solutions, but they are not the same as science. That is not to say that mathematics and statistics cannot be the topic of science. Mathematicians do attempt to explain why and how mathematics and statistics can and do work. Nonetheless, in the structure of science, and in institutional research, such modeling techniques are the products of science applied as tools, and not science. Refutation is the final major building block in the strncture of science as it is practiced today. A hypothesis must be falsifiable or testable in order to make a contribution to knowledge. Popper was the philosopher who brought this concept to the lore (Watkins, 1984). This view became popular as a result of the problems that arose in ruling out alternative explanations for phenomena. Popper realized that it only takes one piece of negative evidence to refute a hypothesis, while no amount of evidence will prove a hypothesis. He argued, "The criterion of the scientific status of a theory is its falsifiability or refutability, or testability" (Casti, 1989, p. 33). Testability is central to the function of institutional research. However, testability has a slightly different meaning for institutional research. The hypotheses and theories tested are the ones associated with the particular institution. Institutional research exists to inform decision making. If there were no alternatives, there would be no reason to make a decision. If there were no way to demonstrate the superiority of one alternative over another, flipping a coin would be adequate decision support. Institutional research and science are not one and the same. That is not necessarily to be perceived as a disadvantage for either one. Pure or basic science is ill suited to the needs for which institutional research was created. The goal of institutional research is to reduce uncertainty in practical decision making. The rigors of science, along with its goal of contributing to knowledge, are rauch too limited for this purpose. Institutional research deals witb more than a single question at a time; instead it deals with an "ecology" (Cleveland, 1988, p. 684) in which a balance must be struck among conflicting factors and factions. Finances, quality of instruction, market share, cultural diversity, student needs, and the influences on and of other social and environmental entities are orten aspects of an issue that taust be addressed in order to info~~ a single decision. Institutional research is more an art than a science. It requires understanding of models, techniques, theories, and laws from other disciplines to inform decision making. However, sorting out what is important to the institntion's viability requires more. lnsights into the institution's character, culture, history, and environment are necessary to understanding the state of the institution's ecology and the nature of the question to be answered. This is especially true when the ambiguity surrounding the question is great. These are things that science generally ignores due to the need to be sufficiently figorous. Science is too narrow to support organizational decision making. KUHN'S PERSPECTIVE: THE SOCIAL STRUCTURE OF SCIENCE The concept of the paradigm (Kuhn, 1962, 1977) provides yet another construct for exploring the links between institutional research and science. A paradigm is a metaphor for understanding the wofld. Some might even equate it with Weltanschauung (world view). A paradigm is a set of shared beliefs about the nature of reality. That set of beliefs determines what is meaningful and important, how the important and meaningful things can be examined, and where answers can be found. Kuhn refers to those who share such a set of beliefs and share in the practice of a scientific specialty as a "scientific community" (1962, pp. 4-6). Kuhn suggests that a scientific community shares a "characteristic set of beliefs and preconceptions" (1962, p. 17). Institutional research has something near to a "scientific community" as defined by Kuhn. The goals of encouraging student success and maintaining institutional viability are shared by those who fulfill the function of institutional research. Many of the tecbniques used in institutional research are shared. The same kinds of problems are the focus of attention of those associated with institutional research. And even though findings of studies may not be generalizable to other institutions, the processes used to arrive at the results may be generalized in that they become the accepted methods of analyzing certain information to inform decisions. Kuhn (1962) also states that members of a scientific community share common educational experience and are socialized to share commitments to certain values. Even though institutional researchers come from diverse educational backgrounds, they usually share some training in the use of statistics, in one or more of the social sciences, and in computer usage. They are socialized to the specialization via professional organizations, conferences, and professional journals. Institutional research is in a pre-paradigm stage; no body of scientific theory controls the kinds of questions that can be asked or the kinds of answers sought. This is nothing for institutional researchers to be embarrassed about. Acdording to Kuhn, "It remains an open question what patts of social science have yet acquired... [universally accepted] paradigms at all" (Kuhn, 1962, p. 15). Yet institutional research does have principles that provide guidance in what is appropriate to study. The principles involve perpetuation of higher education in particular contexts. Even though instimtional researchers share no single paradigm, institutional research has undergone changes on the order of paradigm shifts. Each time a new issue emerges as one of the goals of the institution, institutional research taust adjust to incorporate the concept into investigation of each question posed. Diversity is a prime example. Over time, we have seen a change in the rhetoric used to describe attempts to reduce racial discrimination and increase racial harmony. First, the term was "desegregation." Later it was called "integration." Now it is called diversity. Each change has represented a different focus for instimtional researchers. Each change has resulted in a different treatment of information. Since diversity has become a major goal of institutions, all choices taust be weighed against the possibility of jeopardizing progress toward that goal. CONCLUSION Because institutional research is dependent on the politics of higher education and on the social goals of government and communities, identification of what is relevant will change. This dynamic nature of institutional research demands that the observer shift his or her vision in pattern identification frequently. For that reason, institutional researchers are constrained in their attempts to develop a cohesive body of knowledge. If institutional research were not so dependent on the perceptions of those within a particular institutional context, it might be possible for a discipline to emerge. However, institutional research is a support function for institutions of higher education. An issue is important to a particular institution only when it is recognized as an issue. For the most part, others set the agenda for institutional research. Sometimes it is the government; sometimes it is the school's administration. There is little incentive for this change. Institutional research has many links with science. It makes use of scientific method. It shares elements of science's logical structure. And it is practiced by a particular community that shares some beliefs about what is important to study, and how to study what is important. Nonetheless, institutional research is different from science in meaningful ways, the most important of which is the difference in purpose. Institutional research uses scientific convenfions as tools to achieve its purpose of answering "which" questions. That is not to say the institutional researchers never participate in science; that is not the point. The "theory" of the single institution deterrnines the nature of a study in insfitutional research. The theory of institufions of higher education, that each institution has a mission, goals, strategies, and a set of assumptions about how questions should be approached (whether or not these have been made explicit), is what creates a community of institutional researchers. This discussion has only begun the process of identifying ideal types for institutional research. Refinements of the structure discussed herein, the identification of assumptions, and the recognition of unifying principles in research will contribute to greater understanding of institutional research as a unique disciplined pursuit and assist in the strengthening of the institutional research community.

Research productivity plays a central role in higher education institution development1. Research productivity helps improve teaching quality as there exists a strong nexus between teaching and research effectiveness (Desselle et al., 2018). At the same time, research productivity helps develop knowledge and shape higher education institutions’ reputation and brand names. In actual fact, Porter and Toutkoushian (2006) found that faculty research productivity was positively related to university reputation. In recent years, there has been a competition among Vietnamese higher education institutions to improve their research capability. In 2018, the number of annual international publications in Vietnam reached 10,000 articles. For the year 2018, in comparison with previous year, there was a 34.7% increase in Vietnam publication in general and 41.6% increase in Vietnam universities’ publication in WoS and Scopus database (Chung et al., 2019). Still, research ability of academics in Vietnam universities was inadequate (Nguyen and Klopper, 2019) and research productivity level remained low (Pham and Hayden, 2019). Meanwhile, previous studies showed that active educational policies could promote scientific research and increased university faculty's research capacity. For example, Beerkens (2013) suggested that the effect of management practices on research productivity was consistently positive. According to that author, universities with an suitable management approach not only had higher absolute level of research productivity but also demonstrated faster growth in research productivity. Prendergast et al. (2019) found that communication policy such as enhanced peer mentoring program was effective in improving academic. On the other hand, institutional characteristics such as size, time in operation and location may also affect universities' academic performance, especially research productivity. Although the literature abounds with evidence of effective policies and improvement of appropriate organizational characteristics promoting university research productivity in developed countries, little was done in developing countries like Vietnam. One of a few papers investigated institutional factors affecting university lecturers' research engagement was that of Tien et al. (2019) in which the author employed interpretive qualitative case study approach to examine situation in a higher education institution in the Mekong Delta region of Vietnam. They found that institutional factors affecting lecturers’ research engagement in that university comprised of governmental policies, funding and structure, resources, teaching loads, leadership, and research environment. Moreover, previous studies in Vietnam tend to focus on social science universities in Vietnam such as that of Vuong et al. (2019) rather on science and technology universities. In one of our previous studies (i.e., Nguyen and Nguyen, 2020), we explored the determinants of R&D outcomes in science and technology universities. However, as pointed out in the limitations section of that study, R&D outcomes were analysed merely based on faculty's evaluation and no objective measure of research productivity such as number of international publications was employed. Additionally, our previous research did not include variables reflecting organizational characteristics such as time in operation and location of the universities. Therefore, this current research was conducted to overcome these limitations. We believed that examining factors affecting research productivity at university level rather than from faculty staff level and with a more comprehensive set of variables would provide a richer and distinct insight. This research seeks to contribute to the current literature in the following aspects. First, it was one of the first few studies examining research productivity at the university level of Vietnamese science and technology universities. Second, it examined the effect of institutional policies and organizational characteristics such as size, time in operation and location on university research productivity using an objective measure of university research productivity. Third, the research also examined the link between faculty research outcomes and research productivity at the university level as well as the direct/indirect impact of different variables on university research productivity. Finally, based on the findings, we proposed some policy recommendations to improve university research productivity. The structure of this research article is as follows. The next section presents theoretical framework and research hypotheses. The third section presents research methodology and data. Research results are presented in the fourth section. The final section concludes the article and proposes several policy recommendations. 2. Theoretical framework and research hypotheses 2.1. Research productivity There are several different viewpoints regarding university research productivity. Some authors such as Rivera-Huerta et al. (2011), Kim et al. (2019), Kosyakov and Guskov (2019) focused on the quantity aspects (i.e. research productivity was defined as number of publications or number of patents, etc.) while others such as Kaplan et al. (1997) or Daigle and Arnold (2000) emphasized the importance of research quality (i.e. how research outcomes changed the industry and society, the quality of journal in which the publications were published). Some researchers such as Sahoo et al. (2017) developed research productivity indicators. However, Nygaard and Bahgat (2018) argued that different bibliometric indicators captured different aspects of research performance, including diversity of output and collaboration, which simply reflected different publication practices. In this study, we focused on the quantity aspect of research productivity measured by international publication outcomes. 2.2. University's characteristics and its influence over research productivity Many previous studies examined the relation between university's characteristics and research productivity. Based on these studies, 3 main characteristics were selected for our model, namely size, location, and time in operation. 2.2.1. Size Jordan et al. (1989) found that publishing activity was higher in private institutions and increased with department size at a diminishing rate. Golden and Carstensen (1992) also found that per capita publication increased, up to a point, with department size and is higher at private institutions. This is consistent with the findings of Meador et al. (1992). Gander (1995) used University of Utah data on funded research output to find that as faculty size increased research productivity rose. In another research, Schoenfeld et al. (2015) found that academic affiliation and number of fellows in a program was significantly associated with total number of publications. Similarly, Khan et al. (2017) found that number of fellows, faculty academic title, years in practice, and formal fellowship training had a significant positive correlation with both h-index and number of publications of medical university faculty. Thus, we proposed our first hypothesis: H1: Higher institution of bigger size has higher research productivity than smaller ones. 2.2.2. Time in operation and location In Vietnam, universities located in big cities and have longer time in operation may have advantages in accumulating resources for research and thus have higher research productivity. The reason is that these universities may be more easily reached by talented students and highskilled faculty concentrating in big cities. At the same time, universities in big cities may find it easier to collaborate with big companies locating in those cities. However, previous studies in other countries did not find evidence of this effect. For example, Chan et al. (2001) found that university history and tradition were irrelevant to university research productivity. Thus, we will test another two hypotheses: H2: Higher institution having longer time in operation has higher research productivity than others. H3: Higher institution located in big cities has higher research productivity than others. 2.3. Institutional policy Nguyen et al. (2016) applied a qualitative approach to explore affordances, barriers, and motivations towards research engagement. The findings revealed that most of the respondents were aware of the importance of research, but their research productivity was low because of problems related to institutional factors including research financial support, teaching load, research collaboration, and research policy settings and practices. Institutional policy can be a decisive factors as it influences the time that faculty can spend on research. Smeltzer et al. (2016) examined research productivity of faculty teaching and mentoring doctoral students and found that the strongest predictor was the average number of hours spent on research-related activities, followed by time bought out from teaching and other responsibilities of the faculty role for research. In the following subsections, we will review various literature to form a base for our next hypotheses. 2.3.1. Management policy Management aspect of institutional policies is important for the success of research projects and helps improve quality of research outputs. Management decides the collaboration and sharing of research results as well as intra-university resource allocation. Proper resource allocation and effective cooperation can motivate researchers to implement their research projects and ultimately help improve research results at university level. For example, Uncles (2000) identified three research productivity impediments including inadequate training, sub-optimal concentrations of research activity, and competing commitments. These issues can be dealt with through formal research training and on-the-job practice, developing synergies between research and teaching/consulting, and finding partners to assist in fieldwork and analysis (Uncles, 2000). Similarly, it is found that appropriate university management mechanisms can spur research and positively impact innovation performance (Chanthes, 2012). This can lead to an increase in the productivity of researchers on an individual level due to the positive incentives stemming from openness and efficiency in university governance (Beerkens, 2013). In other words, appropriate university management mechanisms can accelerate research process and participation of scientists in research programs and eventually help increase research productivity at both the individual and university level. Therefore, in this study we propose the following hypotheses: H4: University management policy has positive effect to faculty's research outcomes. H5: University management policy has positive effect to university research productivity. 2.3.2. Human resource policy Carayol and Matt (2006) took account of individual and collective determinants to explain individual's productivity in terms of intensity and quality. According to their study, the intensity and quality of colleagues' research laboratory activities were beneficial for individual research. Nguyen (2016) investigated the extent to which leading universities motivated their academics to improve research performance and found that Vietnamese universities did not have enough powerful human resource policies to encourage academics to do research. Thus, in this research we examine the following 2 hypotheses: H6: University human resource policy has positive effect to faculty's research outcomes. H7: University human resource policy has positive effect to university research productivity. 2.3.3. Communication policy In terms of communication factor, Brodie (2000) suggested that organising annual conference and doctoral colloquium, establishing a journal that publishes high quality research, and establishing website as researchers' networking hub could increase academic research relevance and productivity. Vasileiadou and Vliegenthart (2009) found that communication exchange such as academic meetings was the most important predictor of research productivity. Besancenot et al. (2017) found evidence that size and quality of authors' networks were positively related to productivity. Ho et al. (2017) employed network theory to explore characteristics of a network of 412 Vietnamese distinguished social scientists. High clustering and low density were found to be tied to inefficient expertise dissemination among Vietnamese social scientists, and consequently to low scientific output. Similarly, Valsangkar et al. (2016) found that increased faculty participation in an academic association helped increase scientific impact and productivity among association members. Hafsteinsdottir et al. (2017) found evidence of mentoring's influence on research productivity, career development and other outcomes of postdoctoral. On the other hand, Abramo et al. (2017) provided an in-depth analysis of the relation between the different types of collaboration and research productivity, showing how both were influenced by personal and organizational variables but only intramural and domestic level collaboration had a positive effect on research productivity. On the contrary, Nguyen et al. (2017) found that the vast majority of scientific papers from Vietnam was attributable to international collaboration. As empirical evidence for this factor was inconclusive, we examine the following 2 hypotheses: H8: University communication policy has positive effect to faculty's research outcomes. H9: University communication policy has positive effect to university research productivity. 2.3.4. Infrastructure policy Alrahlah (2016) performed a qualitative study on a group of 21 respondents working in different dental colleges and identified a lack of adequate research facilities as barriers to their research productivity. Similarly, according to Kang et al. (2017), infrastructure aspect such as indoor environmental quality had significant impacts on university faculty's research productivity. Therefore, we examine the following 2 hypotheses: H10: University infrastructure policy has positive effect to faculty's research outcomes. H11: University infrastructure policy has positive effect to university research productivity. 2.4. Financial constraints Impact of financial factors on research productivity was found in various studies. For example, Jacob and Lefgren (2011) estimated a significant impact of receiving a research grant on subsequent publications and citations. Zhang et al. (2017) also found that receipt of government funding was associated with a higher h-index. Similarly, Hottenrott and Lawson (2017) found that research grants were generally associated with higher research outcomes. Zaorsky et al. (2019) found an association between disclosed payment from the industry and increased individual research productivity metrics. Pitt et al. (2017) found that the combination of increased awareness of peers' academic productivity and a weighted lottery financial incentive appeared to be a useful model for stimulating academic productivity in early-career faculty. Therefore, we examine the following 2 hypotheses: H12: Financial constraint has negative effect to faculty's research outcomes. H13: Financial constraint has negative effect university research productivity. In this study, we defined financial constraint as the hinders that universities researchers must face while finding financial resources for their own research. On the other hand, infrastructure policy is the university investment on infrastructure which may affect both research and training and both university and individual's level. 2.5. Faculty's individual ability constraint Individual ability was found to be an important factor affecting research productivity in many previous studies. Huang et al. (2015) found that research productivity increased with departmental academic rank. Shih et al. (2018) affirmed that factors associated with increased academic productivity include attaining higher academic positions, having a longer career length, having greater numbers of research grants, and having MD and PhD degrees. Rubin and Callaghan (2019) found that individuals with higher self-reported levels of entrepreneurial orientation and more eager to apply novel technological research methods have higher levels of research productivity. Jean and Felbaum (2019) identified a positive correlation between the subjects' academic productivity and the ranking of all the institutions throughout their education, training, and current employment. Therefore, we examine the following 2 hypotheses: H14: Faculty's individual ability constraint has negative effect to faculty's research outcomes. H15: Faculty's individual ability constraint has negative effect to university research productivity. H16: Faculty research outcomes has a positive impact on research productivity. We combined all research hypotheses in our research model as depicted in Figure 1. 3. Research methodology 3.1. Measurement scales Measurement constructs reflecting institutional policies used in our research model were developed based on research of Nguyen and Nguyen (2020). Specifically, infrastructure policy was measured by 5 items, management policy was measured by 7 items, human resource policy was measured by 5 items, communication policy was measured by 4 items, financial constraint was measured by 4 items, faculty's individual ability constraint was measured by 9 items, faculty research outcomes was measured by 4 items. These items were measured using questions by the Likert scale with 1 being totally disagree and 5 being totally agree. For example, to measure communication policy, the respondents were asked 4 questions whether they agree or disagree with the following statements (1) Your university maintains good professional network (2) Your university establishes frequent intra-organization academic communication (3) Your university help researchers connect to external sources for R&D activities (4) Your university sufficiently organizes R&D workshops/ symposium/conferences. The items for measurement scales are described in appendix A. Appendix B presents summary statistics of items in measurement constructs. 3.2. Data Main data were collected through questionnaire survey conducted by Association of Vietnam University and Colleges in Vietnam science and technology universities in corporation with Hanoi University of Science and Technology under research project BKA-2017-41 funded by Ministry of Education and Training, Vietnam. Before the research project was carried out, our research proposal including ethical issues was approved by a committee from Ministry of Education and Training, Vietnam specially designated for research project BKA-2017-41 in 2016. We included informed consent in our questionnaire. The participants had the right to refuse to fill in the questionnaire after reading it. In other words, informed consent was obtained from all participants in our research. The questionnaires were distributed by hand to the participants of the National Conference on University autonomy and by mail sent to the target university representatives. At least one representative from the 115 targeted universities sent back a valid questionnaire answer. 585 questionnaire forms were sent to universities in which 468 completed forms were sent back. The number of valid answers was 415. Data about number of student intakes, time in operation and university location was collected from Vietnam Ministry of Education and Training. Data for international publication of the universities were obtained from Scopus database for the year from June 2017 to June 2018. Among 115 targeted universities, 33 universities were private universities, 31 universities were under control of Ministry of Education and Training, the rest was under control of other ministries or provincial governments. The average international publication of the universities was 44. The annual university student intakes ranged from 100 to 7340 with mean value of 2270. The years in operation of the universities under survey ranged from 4 to 117. 3.3. Data analysis method Multivariate data analysis method was used to analyse surveyed data and test proposed hypotheses with structural equation modelling. The reliability of each construct in the model and internal consistency were evaluated by Cronbach's Alpha coefficient value and exploratory factor analysis. We used confirmatory factor analysis method with data conversion to check for construct validity together with common method bias and non-response bias tests. Because the survey was conducted at individual level (i.e., information was collected from university representatives) while the aim of the study was to evaluate the effect of policies and characteristics on research productivity at the university level, we transformed the data from individual level into university level by the following formula: xi ¼1 nX n 1 xij (1) In which: xi is the value of item ith in the construct at university level. xij is the value of the item at university i n is the number of respondents in university i. Number of student intake (size), number of scientific papers published in Scopus index (research productivity) and time in operation vary greatly among the surveyed universities. Therefore, we used natural logarithm of each variable to reduce heterogeneity in the estimation. As mentioned earlier, this research was extended from one of our previous study (Nguyen and Nguyen, 2020). For comparability purpose, we applied a similar analytical method. However, there were some significant differences between the methods being applied in the 2 models. In the research model of our previous paper, we applied individual approach (i.e., from faculty's perspective). Thus, we had 415 observations representing opinions of 415 faculty researchers. However, in the research model of this article, we applied institutional approach and aggregated the opinions of faculty of the same university and transformed the data from individual level into university level using Eq. (1). Thus, we had only 115 observations representing 115 universities in the model. In addition, in this study, we focused on the determinants of various factors on university research productivity which was proxied by total number of international publications. This was different from our previous research (Nguyen and Nguyen, 2020) in which the dependent variable was faculty research outcomes. In this research, faculty's own evaluation about their research outcomes was used as an intermediary variable which also affected university research productivity. Direct and indirect impact of various variables on university research productivity were also examined in this research. 4. Research results 4.1. Reliability and validity The analysis results indicated that all constructs in the model reached satisfied reliability conditions. Specifically, Cronbach's alpha coefficients were all larger than 0.7; KMO was larger than 0.5 with p-value < 0.05. Total variance explained was 64% which was much higher than the threshold level of 50%. Factor loadings of each items in the constructs were greater than 0.5 (Table 1). The findings of confirmatory factor analysis showed that our research model fit the actual data, specifically Chi – square/df ¼ 1.578 was less than 3, CFI ¼ 0.914, TLI ¼ 0.903, IFI ¼ 0.915 were all greater than 0.9, and RMSEA ¼ 0.071 was less than 0.08 (Table 1). The factor loadings of each item in the constructs were greater than 0.5 showing that the constructs in the model reached convergent validity (Hair et al., 2018). The composite reliability coefficients were greater than 0.7 and average variance extracted was greater than 50% indicating that the model constructs were reliable. The value of the square root of AVEs was greater than any correlation coefficients in the model which showed that the constructs reached discriminant validity (Table 2). 4.2. Common method bias and non-response bias Because common method bias may influence the true relationship between constructs (Podsakoff et al., 2003), we designed the constructs to avoid ambiguous items and control for acquiescence and dis-acquiescence biases. Harman test results indicated that when fixed to a unique factor of all items, the total variance explained was 24.868% which was much smaller than 50%. This meaned that common method bias did not affect our study results. To examine non-response bias, we used t-test to compare early respondents and late respondents divided at a ratio of 70:30 (Armstrong and Overton, 2018). The findings found no difference between the two groups implying that non-response bias was not a concern in our research. 4.3. Hypothesis testing We used structural equation modelling to test our research hypotheses. The results showed that our model fit with actual data: Chi – square/ df ¼ 1.671 was less than 3; CFI ¼ 0.905, IFI ¼ 0.907 were all greater than 0.9, and RMSEA ¼ 0.077 was less than 0.08. The estimation results were shown in Figure 2. Estimation results with standardized coefficients indicated that eight hypotheses were accepted, including: (H1) Size of university is positively associated with research productivity (β ¼ 0.34, p-value < 0.05); (H2) Time in operation of university is positively associated with research productivity (β ¼ 0.23, p-value < 0.05); (H3) Location (in a big city) is positively associated with research productivity (β ¼ 0.13, p-value < 0.05); (H5) Management policy has a positive impact on research productivity (β ¼ 0.35, p-value < 0.05); (H6) Human resource policy has a positive impact on faculty research outcomes (β ¼ 0.21, p-value < 0.05); (H11) Infrastructure policy has a positive impact on research productivity (β ¼ 0.34, p-value < 0.05); (H16) Faculty research outcomes has a positive impact on research productivity (β ¼ 0.15, p-value < 0.05); (H8) Communication policy has a negative impact on faculty research outcomes (β ¼ -0.21, p-value < 0.05). Other hypotheses in the model were rejected. In other words, the results supported hypotheses H1, H2, H3, H5, H6, H11, H16 and rejected hypotheses H4, H7, H9, H10, H12, H13, H14 and H15 (Figure 2). We found an unexpected relationship contrary to hypothesis H8. To assess the effect of each construct on the research productivity of university we used total effect coefficient. The findings showed that the largest effect was MAN (γ ¼ 0.349), and smallest was HUM (γ ¼ 0.029) (Table 3). 5. Conclusions and policy recommendations 5.1. Summary of key findings The research results helped answer our research questions about the impact of institutional policies and university characteristics on university research results and research productivity. The study also verified the influence mechanism of research development support policies and university characteristics on faculty research results and research productivity at institutional level. In general, the results were consistent with previous studies (Akl et al., 2012). However, there were certain differences stemming from the context and specific characteristics of Vietnamese universities. The study detected a significant impact of university characteristics on university research productivity. Specifically, university size proxied by the number of enrolments had a strong influence on universities' research productivity (β ¼ 0.34). Time in operation also affected research productivity (β ¼ 0.23). The authors also found a significant but relatively weaker influence of location (β ¼ 0.16) on university research productivity. These results were quite consistent with previous studies showing that the university's published scientific output was influenced by organizational characteristics such as size, location and years in operation (Schoenfeld et al., 2015). The results supported the argument that a larger university would have more faculty and researchers involved in research activities and consequently lead to a greater number of scientific development policy on research outcomes at the faculty level. This implied that human resource development policies such as attracting talented and active faculty were crucial to develop faculty research activities and indirectly increase university research productivity. This result was supported by previous studies in which university human resource development had a positive impact on research activities at both individual and institutional level (Carayol and Matt, 2006; Nguyen, 2016). In terms of communication policies, contrary to our expectation, this factor had a direct negative impact on faculty research outcomes and therefore had a negative impact on university research productivity. This needs to be explained in the context of this study. The research productivity in this research was judged based on the number of articles published in the Scopus database and did not consider the articles published in domestic journals. Therefore, research productivity here was only evaluated according to international standards and thus could be overlooked in other scientific activities. Another fact that should be taken into consideration was that the request for scientific funding from scientific funds in Vietnam did not require international publication until recently. For example, the requirements for research project funding of Ministry of Science and Technology before 2018 did not list international publication as mandatory requirement. Therefore, there was a tendency for researchers to publish in domestic journals to speed up scientific funding. This may be the reason for the increased level of interpersonal engagement for research projects that did not promote international publication where project managers tend to opt for more domestic journals instead of international journals. Consequently, improvement in communication policies did not help increase research productivity according to international standards. We also found positive direct effects of management policies and infrastructure development policies on university research productivity. Data analysis showed that for science and technology universities, policies for large capital investment or research activity support such as laboratory construction, experiment equipment installation significantly affected the research capacity of university lecturers and researchers. This result reinforced the arguments from previous studies that developing research, especially research in the fields of science and technology, required large investments in research infrastructure (Alrahlah, 2016; Kang et al., 2017). The logic of creating high research productivity was to invest in the necessary conditions for research activities (i.e., infrastructure) and to put in place an incentive-based management mechanism encouraging research activities. In comparison with our previous study (Nguyen and Nguyen, 2020), there are some different results which reflected the differences between faculty's perspective approach and the institutional level approach. Firstly, our previous study showed that financial factors were one of the two variables that significantly affected faculty's research outcomes. However, in this research, it was shown that management factor rather than financial factor significantly affect research productivity at university level. It may be because faculty tended to overemphasize financial factor as their activities can be directly and individually hindered by financial constraints. The authors believed that financial support affected research motivation especially from faculty's perspective. However, if financial support was too small to influence the research efforts of university faculty members, it may not affect their research or non-research decisions. This may be especially true if it comes to international standard research such as Scopus journal publication which requires lots of time and efforts. The fact that the financial support for research activities at Vietnamese universities was still modest may be the reason for the negligible impact of financial support on university research productivity at institutional level. Secondly, in this study, human resource policy had significant indirect impact on university research productivity at institutional level while this variable did not significantly affect faculty research outcomes in our previous study. This showed that examining factors affecting research productivity at individual and university levels generated different pictures. Because there was a strong connection between faculty's research outcomes based on their own evaluation and university research productivity, if the universities want to quickly boost the number of international publications, they should focus on infrastructure and management policies but at the same time should not neglect human resource. Lastly, it is worth noting that while size of the university had no significant impact on faculty research outcomes in our previous research, this current study showed significant impacts of this variable and other institutional characteristics variables on university research productivity. This might be explained by the fact that university size and fame tended to be overlooked by faculty when subjectively evaluating their individual research outcomes. However, these variable and other institutional characteristics should not be left aside when examining research productivity at university level. 5.2. Implications for policy Based on the above-mentioned research results, we proposed some suggestions for improving university research productivity. Firstly, scientific funding management mechanism needs to be vigorously revised in direction of simplifying the process of reviewing and approving research grants, reducing administrative procedures, and increasing commitments in terms of international publications outputs. Secondly, universities should be encouraged to invest in building infrastructure for research and training activities, attracting good researchers to work at the university, forming strong research project teams towards increase research productivity. Universities should consider investing in infrastructure and developing human resources for research as a priority to increase research productivity and improve universities’ position in academic rankings. Universities need to strengthen their linkages with enterprises and attract financial sources from enterprises in forms of scientific grants, reducing the dependence on government research funds. Thirdly, the location of new university campuses should be carefully assessed so that the university can attract students to generate financial sources for the universities’ activities and promote scientific research. New campuses should be located near economic centres and large cities to attract lecturers and learners. Fourth, universities should consider policies that motivate their academics in both teaching and researching, helping them to break out of their ability constraint. 5.3. Research limitations and implications for future research This research contains some inherent limitations. First, due to its design, the research could not examine faculty characteristics which might affect research productivity such as gender, age, degree, rank as mentioned by Paik et al. (2014), Vuong et al. (2017), Adib et al. (2018) or discipline and years of work experience as mentioned by Shih et al. (2018), Nafukho et al. (2019) and Mueller et al. (2016). Examining these characteristics might solve the unanswered questions in our research such as why older universities have more advantages in terms of history and fame than younger universities. Second, the research used international publication in Scopus database as a simple measure of research productivity. This measure might not reflect all research productivity aspects. Third, another important limitation of the research was that the variables measuring institutional policies were based on the opinions of university representatives, rather than on objective data. Therefore, one person's view on, for example, "sufficient faculty members", might not be the same, even if they were at the same university. Last, the data just covered science and technology universities which accounted for about one third of all universities in Vietnam. Future research should seek remedies for these limitations to depict a more complete image of factors affecting research productivity in Vietnam.

Journal impact factor (JIF) introduced by the Institute of Scientific Information (ISI) by means of the journal citations report (JCR) has a long tradition as an impact factor (IF) indicator for scholarly research output. IF and its variants have been introduced and displayed on JCR site (www.webofknoweldge.com) as a measure of research quality or impact of journals. In general research performance evaluation (RPE) practices, IF has become a ‘‘chief quantitative measure of the quality of a journal, its research papers, the researchers who wrote those papers and even the institution they work in’’ (Amin and Mabe 2000 p. 2) however, it cannot be used as a direct measure of quality (Bornmann et al. 2011). Although, IF remains the primary criteria when it comes to assessing the quality of journals and authors (Raj and Zainab 2012), it should not be used as a sole measure of a journal rank (Bornmann et al. 2011). To overcome the limitations of IF, researchers suggested that it could be used with new alternative tools (Braun et al. 2006; Borrnmann et al. 2011; Prathap 2011; Yin 2011) or as a measure of research quality or impact of journals (Braun et al. 2006). The notion of journal’s h-index was introduced by Braun et al. (2005), who later found that h-index could be used as a measure of research quality or impact of a journal (Braun et al. 2006). After the introduction of h-index, a number of studies made a comparative analysis of both impact indices (h-index and IF) and their variants. These indices are easily comprehensible (Leydesdorff 2009) and have received worldwide recognition. However, the previous studies, as reviewed in the subsequent paragraphs were concerned with the evaluation of journal’s h-index to journal related indices (JRI). Mingers et al. (2012) examined journal level h-index against impact factor (IF), 5-year impact factor IF(5y) and peer judgment for management journals. They preferred journal h-index to IF because of the former’s selective time frame and the formulaic problem. Another journal study in the field of marketing was carried out by Moussa and Touzani (2010) using Google-Scholar (GS) as the source data. They used a variant of h-index, i.e. the hg-index along with 2 and 5 years IF. There was a substantial agreement found ([0.85) between IF5y and the hg-index ranking. They suggested hg-index as an alternative to the GS-based journals. Soutar and Murphy (2009) studied 40 marketing journals and ranked them according to IF and h-index, and compared their list with the Australian journal ranking. They suggested these indices as the basis for moving some journals up and other journals down in the list. Their study supported the use of GS as an alternative way to measure citations in the field of marketing. Harzing and Van der Wal (2009) compared the h-index generated from the citation in GS with the impact factors computed from the Web of Science (WoSTM) of over 800 business and management journals. They concluded that the GS h-index provided a more accurate and comprehensive measure of journal impact, as other studies indicated that WoSTM did not have a good coverage across all subject areas (Abrizah et al. 2013; Tahira et al. 2012; Gavel and Iselid 2008; Hicks and Wang 2011; Mikki 2010). A comparative analysis of IF and h-index was carried out by Bador and Lafouge (2010) on pharmacology and psychiatry journals from JCR with 2-year publications. The journals correlation coefficient between IF and h-index was high. They inferred that IF and h-index could be totally corresponding when analyzing journals of the similar scientific subject. Bornmann et al. (2009) studied the journal’s h-index of twenty organic chemistry journals from WoSTM database for 2 year-time span. They analyzed a number of impact indicators including IF and journal’s h-index and its variants g index, h2 index, A, and R index, and found a high degree of correlation between the various measures. Combining the h-index with another measure in JCR i.e. the Eigenfactor, Yin et al. (2010) analyzed top journals in the field of science and engineering using data from WoSTM. They provided graphical plots of the relative positions of these journals and concluded that the two indices measured slightly different characteristics. In another study, Yin (2011) analyzed 20 top journals in chemical engineering also using data from WoSTM comparing journal h-index and Eigenfactor score. The researcher hypothesized ‘‘that the combination of complementary journal indicators could provide a simple, flexible and practical alternative approach for evaluating scientific journals’’ (p. 2). There is a positive correlation, although not strong among these indices and this indicates that the two indices can also be combined to complement each other. Yin suggested authors to get their research work published in high Eigenfactor scores journals. The objective of past studies was to evaluate a journal’s h-index validity and reliability with respect to other JRI. Most of these studies considered journal’s h-index as a contrasted measure with JCR indices such as IF, IF(5Y), and Eigenfactor score. These studies are meaningful to understand the properties of newly introduced indices (Tahira et al. 2014) and potential use of journal’s h-index as a complement aid with IF and its variants (Bador and Lafouge 2010; Bornmann et al. 2012; Yin 2011) or, as a supplement (Braun et al. 2006). These all studies are limited in various ways, especially concerning the relation of JRI and institutional level h-index (IHI). To have a further evidence of the reliability of h-index at the institutional level communicated in Tahira et al. (2015) and a set of impact indices (including total citation and JRI), we have hypothesized that IHI is a potential index for RPE that can be used as a complement or as a supplement for RPE at the institution level. Objectives and method The objective of the research is to evaluate the reliability of Institutional level h-index (IHI) with respect to other journal related indices (JRI). We examined the intercorrelation by exploratory factor analysis (EFA). In addition, we have explored the functional relationship by regression analysis and checked the correlation among IHI and a set of impact indices to evaluate its reliability. The empirical part of this study focuses on one non- Western country, Malaysia, which has a developed and well-defined scholarly publishing industry based in its universities. Research productivity, citations record, and institutional journal data of twelve Malaysian universities were retrieved from WoSTM and JCR 2011. The universities having at least 50 publications (journal articles) during the past 10 years were selected. A total of eleven indices were used for the present study at the meso level. The indices are total publications (TP), total citations (TC) citation per publications (CPP), Institutional h-index (IHI), IF, cumulative impact factor (CIF), 5 years journal impact factor IF(5y), cumulative 5 years journal impact factor CIF(5y), average impact factor (AIF), median impact factor (MIF), immediacy-index (Imm-index) and Eigenfactor score (EF).Their definitions and the acronyms used are described in Table 1. To get a meaningful evaluation, we used a set of WoSTM engineering journals (1381 journals with 68 unique titles) published by 12 Malaysian universities with a wider horizon of 10 years (2001–2010) under specified nine categories. Malaysia’s scholarly publication strength lies in the field of engineering and energy. The search term used was ‘‘Malaysia’’ in ‘‘address’’, limited to document type research article and reviews only, and refined by nine engineering research categories. These nine engineering categories are: electrical and electronic, manufacturing, biomedical, industrial, civil, chemical, mechanical, environmental and multidisciplinary. An overview of data Data were suffered from affiliation problem, change of journal title and abbreviation of a journal name. The data were checked manually for publications, citations, institutional affiliation, and journal name change or emergence cases. Almost all journals in our data set had impact-factor, except for 22 articles published in six non-IF journals. These data were included in the journal list for analysis purpose. All records were retrieved in a spreadsheet file, and IBM statistical product and service solutions was used for statistical analysis. Table 2 presents the university-wise journal contribution in the journals. The publication share of research-intensive universities (RU) was 66 % (908) while the non-RU status universities shared 34 % (473) of the total journals. The RU universities are more bound to published in IF journals to get more research funding. These five universities receive a big amount of budget as research grants for research and development (R&D) purposes, and have to face the publication pressure to build an international reputation in scholarly publication, and this is especially prevalent among the Asian countries (Leydesdorff 2009). Since their research university status in 2007, the country’s research activities and outputs have shown a tremendous increase, which proves that higher learning institutions in Malaysia can be at par with reputable universities in the world. The first five public universities (RU) have published in 150–200 journals. Comparatively, the non-RUs had fewer publications and published in less than 100 journals. The average number of journals publishing articles affiliated to RU and non-RU universities is 182 and 68 respectively. The statistical methodology of exploratory factor analysis (EFA) was used to examine for ‘‘latent associations present in a set of observed variables, and reduce the dimensionality of the data to a few representative factors’’ (Schreiber et al. 2012 p. 349). It is mainly used to identify a smaller set of salient variables from a larger set and to explore the underlying dimensions or factors that explain the correlations among a set of variables (Conway and Huffcutt 2003). Analysis and findings In a tie with the problem, this section proceeds accordingly with descriptive statistics, data normality and EFA for the set of indices used as presented in ‘‘Appendix’’. Descriptive statistics along with Skewness and Kurtosis are presented in Table 3. The Skewness and Kurtosis are valid tests to find the normality of data. Their values show a normal distribution of data. The results of the normality test based on raw data (excluding outliers) are reported in Table 4. Keeping in view the requirement of EFA statistical application, we have used two other options as well. We also examined the relation between the raw data (X), logarithmically transformed shifted (ln(x ? 1) and square root transformation data (Hx). Table 5 shows a better Kaiser–Meyer–Olkin (KMO) results and a slightly better explained variance for log data. For this reason, we found that the logarithmic transformed data is more adequate for EFA. Bornmann et al. (2008, 2009) used a cut-off threshold[0.6 for extraction loading factors while Schreiber et al. (2012) fixed it at[0.685 for Varimax rotation. Schreiber et al. (2012) argued that small sample size for EFA could produce reliable results. Quite a few factors and high communalities are in favour of small sample sizes (Preacher and MacCallum, 2002). Furthermore, to measure a sampling adequacy, a specific test Kaiser–Meyer–Olkin (KMO) of value[0.5 is acceptable (Kaiser 1974). KMO value of the present data sample is[0.5 (Table 5) with high communalities ([0.85) (Table 6). Based on KMO values and variance explained (Tables 5, 7), we finally utilized logarithmically transformed data. We identified two unknown factors through Eigenvalues ([1) via variance explained. This is evident that EFA can be used and is appropriate for our formulated problem and data set. Initially, we considered eleven indices, TP, TC, IHI and 8 of JRI (JIF, CIF, AIF, MIF, IF (5y), CIF(5Y), Imm-Index, and EF). This set of indices produced inadequate results for EFA. After omitting the TP, we applied EFA to TC, IHI, and 8 JRI. Table 5 reports the results of KMO values of the transformed data for the appropriateness of factor analysis. Table 6 reveals the results of communalities for 3 EFA models that are the ‘‘variance in observed variables accounted for by a common factor’’ (Hatcher 1994). Table 7 provides initial Eigenvalues[1 and indicates that the total variance explained by first two factors is 75 and 17 % while cumulative variance explained by both factors is 91 %. Scree plot (Fig. 1) and component matrix (Table 8) illustrate that the set of indices clearly loads on two extracted factors. Rotated Component Matrix (Table 9) for EFA model shows that the indices have substantial loading on two established factors. It indicates the loading of two institutional ‘impact of the productive core indices’ (TC and IHI) and six others JRI have high loading ([0.90) and a slightly less for EF ([0.891). AIF and MIF both have substantially high loading on the second factor[0.9. MIF is more accurate measure than the average value, due to the impact factor’s skewed distribution (Costas and Bordons 2007). IF and CIF and IF(5y) and CIF(5y) require 2 and 5 years time span with different strengths of productivity. EF is another index based on 5-year data excluding journal self-citation to rate the total importance of the journal. Journals generating a higher impact on the field have larger Eigenfactor scores (Bergstrom 2007). Yin et al. (2010) has found that EF improves upon JIF and ‘‘somewhat robust indicators of quality and prestige of the journal due the inclusion of 5 year’s data, exclusion of journal self-citations’’ (p. 3). Rather a high journal EF depicts producing of high-impact scientific findings in a specific area (Yin et al. 2010). IF(5y) indicates the speed with which citations to a specific journal appears in the published literature. Immediacy index that is based on 1-year data shows the same value as CIF on the first factor. They both require a different strength of data. Surprisingly they all loaded on the same factor along with IHI. Functional relationships and predictive values of indices We tested the regression relation of JRI indices uploaded with IHI on the first component. To examine the functional relationship, if any, and their predictive value, we seek IHI model fit and power law relationship (Fig. 2a–h). The results reveal that IHI has a good model fit with IF (R2 = 0.779), IF(5y) (R2 = 0.792) and with cumulative indices, and this fit is improved for CIF (R2 = 0.867) and CIF(5y) (R2 = 0.831). This relation is also found good for both immediacy index and EF (R2 = 0.777 and 0.712 respectively) while not good for the cases of AIF and MIF (R2 = 0.16 and .0017). In this case CIF is a strong predictor for IHI followed by CIF(5Y). The power regression trends between IHI and IF, CIF, IF(5Y), CIF(5Y) are IHI = 0.32IF0.75, 0.56CIF0.53, 0.28IF (5y) 0.76 and 0.53CIF (5y) 0.54 respectively. The relation between IHI and new JRI are found to be 1.39 imm-index 0.720 and 11.54EF0.67. The above values express that the square of IHI is approximately the multiple of CIF and CIF(5Y). At the institutional level, the component analysis and functional relationship of cumulative indices resulted in a much stronger association with IHI (Table 10). Correlation analysis between IHI and JRI To seek the correlation among our set of indices between IHI, TC, and JRI indices, we checked spearman two tail correlation. TC, IF, CIF, IF(5y) CIF(5y), are correlated with all impact indices except AIF and MIF. AIF is correlated with imm-index and EF only. AIF and MIF have no correlation with each other. MIF has no relationship with any JRI, TC, Imm-index and EF. EF is also correlated with all indices except MIF. Discussions and conclusions The caveats of h-index, JIF, and traditional metrics have been discussed in the abundant literature. Previous studies are meaningful to understand the properties of newly introduced indices and potential use of journal h-index as a complement aid with IF and its variants (Bador and Lafouge 2010; Bornmann et al. 2012; Yin 2011) or, as a supplement (Braun et al. 2006). The present study describes the case of Malaysian engineering research by applying the scientometric approach, method and techniques for research performance evaluation (RPE). Based on the 10 years data analysis from WoSTM, we applied a set of comparatively new indices. To achieve the research objectives, empirical analyses were carried out, and the hypothesis that ‘IHI is a potential index for RPE that can be used to complement or as a supplement along with JRI for RPE at the institution level’ was examined statistically. The major findings of the study demonstrate that there seems to be an increasing trend to get published in IF journals. A steady increase of IF publications is observed from 2001 in the Malaysians scientific productivity of engineering research. The desire to publish in IF WoSTM recognized publications is reinforced by the Malaysian Research Assessment (MyRA) exercise, which requires institutions to publish articles that are indexed in global citation databases such as WoSTM and Scopus. This is due to Malaysian Ministry of Education’s policies towards research and publications during two five years plans (2001–2005; 2006–2010). RU status universities shared 68 and 74 % of total publications and citations respectively. These universities have published in 66 % of the total journals involved in this study. Overall, the RU universities lead in positioning order with the application of indices. USM is an exceptional case and remained in position one with respect to almost all indicators while others show a noteworthy change in their positioning order except for UNITEN. Often used metric C (as total impact indicator), imm-index and the EF (as prestige indicator) have a high association with IHI. This establishes the property of h-index as prestige impact measure of scientific productivity. This index appears as a useful yardstick, because of good functional relationship with C and P and has shown some discriminatory power for the ranking purpose. The EFA suggests the same distinguishing be haviour of IHI and total citation. IHI has a stronger functional relation with cumulative IF for 2 and 5 years. Comparatively cumulative IF 2 and 5 years have better model fit than the IF, IF(5y), imm-index and EF. CIF is a strong predictor for IHI followed by CIF(5y). The results of correlation matrix indicate that total citations, institutional level h-index and journal related indices except AIF and MIF have a significant correlation. These two indices are not correlated with each other. MIF shows no correlation with any indices while AIF is correlated with imm-index and EF. EF is correlated with all except MIF. The findings put forward a better understanding of considering the new impact metric for RPE at the meso level. The Malaysian engineering institutional case indicates that h-index and others metric have not only strong functional association for institutional cumulative journal indices but also with total institutional citation data and shows correlation as well. However, rotation loading of variance explained for two components yield about 73 % for its first component and 18 % for the second component. Therefore, findings are based within the limitations of the statistical analysis. Publishing in high-quality IF journals is important if a country is to realize its ambition to have its universities amongst the top rated universities in the world. This is not peculiar to Malaysia. The Ministry of Education Malaysia is targeting one research university in the country to be in Asia’s top 25, and two in the world top 100 best universities by 2025 (Malaysia, 2015). Other countries also place a high emphasis on publishing in IF journals and would want to be ranked as top world universities, even if they are not always explicit in saying so. Given the significant number of papers that have now been published by Malaysian institutions (56,571 in Web of Science, Essential Science Indicators, Web of Science (2015), there is an opportunity to carry out further analysis. It would be interesting, for example, to provide analysis at a discipline level to get a feeling for the strengths of the institution at a lower level. It would also be informative to consider other normalization measures to ascertain if they provide a better correlation with the MyRA ranking.

I have been a devoted fan of college and university offices of institutional research for longer than I care to remember because data they collect has been the basis of much of my research and thus my professional reputation. I also had the great pleasure of serving as the vice president at Cornell University to whom the office of institutional research reported during part of the 1990s. Much to the surprise of our President and Provost, neither of \*School of Industrial and Labor Relations, Cornell University. \*\*Address correspondence to: Ronald G. Ehrenberg, School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14853-3901, USA. E-mail: rge2@cornell.edu whom was data driven, institutional researchers and I conducted two studies during my tenure that directly influenced decisions at Cornell and also led to two published papers, which I will discuss below. Since my return to the faculty in 1998 and my founding of the Cornell Higher Education Research Institute (CHERI), I have continued to extensively use institutional researchers’ data. Indeed, earlier this year I published a paper that made use of data from one data exchange, which I agreed not to identify, on graduate assistant stipends and working conditions, to address some issues relating to the collective bargaining for graduate assistants at public universities. Many institutional research offices have been kind enough to respond to the institutional surveys that CHERI undertakes each year, including our most recent survey requesting permission to access the information on faculty salaries by discipline that a set of universities annually report to the Oklahoma State University Faculty Salary Survey. So put simply, I owe institutional researchers a great debt of gratitude. This paper will provide a quick tour of some of the research that others and I have conducted using data sets collected by offices of institutional research to provide an ‘‘outsider’s’’ view of how valuable the data collected is to researchers and in framing institutional policies. Given the upcoming assessment of graduate programs that will be undertaken by the National Research Council, I will spend most of the paper discussing research relating to graduate education.1 I will conclude with a few comments on how the usefulness of institutional researchers to universities can be improved and stress that institutional researchers and administrators would be wise to involve more faculty members in research that aids in institutional decision making and the formulation of public policy towards higher education. OPTIMAL FINANCIAL AID POLICIES FOR A SELECTIVE UNIVERSITY My first paper relating to the economics of higher education was published in 1984.2 It presented a utility maximizing model that showed what information a selective college or university needed to determine the size of the financial aid package that it should optimally offer to different admitted undergraduate applicants. This paper, which provided the intellectual underpinnings for what is now known as ‘‘preferential packaging’’, required me to estimate a model of admitted applicants’ propensities to accept offers of admission—in particular to determine how sensitive different types of admitted applicants’ acceptance decisions were to the financial aid packages they were offered. To estimate these models required me to merge four different types of data together: 1. Information on admitted applicants characteristics that was ‘‘owned’’ by Cornell’s Office of Admissions. 2. Information on admitted applicants’ families’ financial situations and the financial aid offer that was made to each admitted applicant—these data were owned by Cornell’s Undergraduate Financial Aid Office (remember admissions decisions are need blind at Ivy League universities) 3. Information from an Admitted Student Questionnaire that had been administered to all admitted applicants by Cornell’s Office of Institutional Research. This provided information on whether an admitted applicant was enrolling at Cornell or another institution, the applicant’s second choice institution if she was enrolling at Cornell and where she was enrolling if she was enrolling elsewhere, and the financial aid package that the applicant was offered at the other institution). 4. Information on characteristics (SAT scores, distance from the applicant’ home) of the institution at which the individual was enrolling if the individual was going elsewhere or of the individual’s best alternative institution if the individual was enrolling at Cornell. These data were obtained from published volumes and a geographic information system (inasmuch as this paper was written in the early 80s, the latter consisted of a map of the United States and a ruler). While few people talked about the need for data warehouses in the early 1980s, note that the three Cornell databases that provided individual level data that were used in the study were ‘‘owned’’ by different offices of the university and a crosswalk had to be developed to merge the three data sets together. As will be come clear throughout the paper, although offices of institutional research do not need to ‘‘own’’ all of the data bases that universities maintain, they need to have access to many of them. THE AAU/AGS PROJECT FOR RESEARCH ON DOCTORAL EDUCATION In 1988, the Association of American Universities (AAU) and the Association of Graduate Schools (AGS) initiated a Project for Research on Doctoral Education. Information was collected on all applicants to Ph.D. programs in 10 fields at participating AAU institutions. Longitudinal data sets were created that included information on admissions and enrollment decisions, types of financial aid offered to students at each institution at which they were accepted and, once a student was enrolled in a program, information on the student’s progress through the program to dropout or degree. Ultimately, the project was discontinued because of the difficulty participating institutions experienced in collecting longitudinal data on enrolled students on an ongoing basis. However, the usefulness of these data was enormous. Departments used the data to learn exactly who their competitors for Ph.D. students actually were. Perhaps, more important, the data could be used for research purposes that proved important for both individual institutions and our nation’s system of graduate education. Cornell University participated in the project and the information that Cornell’s Graduate School reported to the AAU/AGS project on continuing graduate students’ status and form of financial support each year had been collected (and archived in paper form) for decades by our Graduate School. After tediously converting these data to electronic form, a graduate student and I were able to use these data to estimate competing risk duration models of graduate students’ propensities to drop out of their programs and their times-to-degree.3 In a paper published in 1995, we showed that after controlling for measures of student ability, students who received fellowship or research assistant support had much higher probabilities of completing their programs and slightly shorter times-to-degree than students who were supported primarily by teaching assistantships or who were self-supported.4 This research supported efforts by the Cornell Graduate School to obtain more resources from the central administration for first-year Ph.D. fellowships. One issue that our nation’s research universities persistently must confront is the claim that our enrollment of foreign Ph.D. students is displacing American students who otherwise might gain admission to our Ph.D. programs. Concern has been expressed, in particular, that foreign students are displacing underrepresented minority students and that this contributes to the continued under representation of minorities among the Ph.D. population. In a wonderful paper published in 1997, Richard Attiyeh, long-time Graduate Dean at the University of California—San Diego, and his son (a Ph.D. economist) used data from the AAU/AGS project to address this claim.5 They estimated models of the probability of students being admitted to Ph.D. study at each institution. Holding constant measures of academic ability, they found that foreign students were ‘‘discriminated against’’ and that underrepresented minority students received preferential treatment in the Ph.D. admissions process. While this study does not imply that American universities are doing enough to attract underrepresented minority Ph.D. students, it did show that programs at AAU institutions were making efforts to increase the flow of minority Ph.Ds. DOES FEDERAL SUPPORT FOR GRADUATE STUDENTS DISPLACE INSTITUTIONAL SUPPORT Much of the data that offices of institutional research compile is done for required reports to government agencies. What many people are not aware of is that these institutional level data are often of great use to researchers conducting policy research. For example, in a paper published in 1993, two graduate students and I used data from a number of IPEDs surveys to estimate the extent to which universities alter their internal allocation of funds to support graduate students in response to changes in external funding that the universities receive to support graduate students.6 We found that increases in federal support for graduate students are associated with a decline (but less than proportionate) in institutional support for graduate students, that the decline is greatest at the top research universities (which have a strong sense of the size they want their graduate programs to be) but that even here the decline is modest, and that when external support for graduate students increases in one field, institutions often divert a share of their internal funds for graduate students’ support from that field to other fields. While some ‘‘diversion’’ of internal support to areas other than the ones that funders are supporting takes place, the magnitude of this effect is not large. Hence the diversion is not an issue that policy makers need worry about. THE 1995 NRC RATINGS OF DOCTORAL PROGRAMS The 1995 NRC Ratings of Doctoral Programs was published while I was supervising Cornell’s Office of Institutional Research.7 Unlike previous doctoral program ratings that had been undertaken, data were also presented in the published volume on objective measures of each program (faculty size, publications, citations, receipt of research grants etc.). While faculty raters of programs were not provided this information at the time they made their ratings, it dawned on me that we could act ‘‘as if’’ they knew this information and thus could estimate models of how their subjective ratings were related to objective measures of faculty size and productivity. We could then use these estimates to understand why the programs at Cornell that were not rated in the top 10 in the nation in their fields were not rated at that level.8 Was it because they were too small—ratings are known to correlate with faculty size? Or was it because the faculty members in the program were simply not as productive as their colleagues in top departments—as measured by things like publications per faculty member and citations per faculty member? The methodology that Peter Hurst and I developed, and then applied to data from Cornell University, was published in a paper in 1998 to illustrate to researchers at other institutions how they could similarly apply our methodology to evaluate the reasons for any of their departments’ failures to be ranked highly.9 Without going into the details of the methodology, as I have described elsewhere, our approach had an impact on at least two graduate fields at Cornell.10 Our sociology department was one of the lowest ranked social science departments at the university and the dean of the College of Arts and Sciences was contemplating withdrawing support from it. However, our analyses suggested that faculty in the department were every bit as productive as faculty at top 10 sociology departments in the nation; their low ranking was primarily due to their small size. As a result, the dean decided to increase, rather than decrease, funding for the department. Today it is a much-improved department, perhaps the best social science department at the university. In contrast, the size of our faculty was seen not to be the major cause of our relatively poor rankings in biology and, after the mandatory period in which faculty members blame the university administration for all the bad things that have happened to them, our biology faculty got together and constructively planned ways to improve what they were doing. This research illustrates that institutional researchers should be opportunistic. They should take advantage of data that others have collected and the methodologies that others have developed and applied to help guide decision-making at their universities. CONFRONTING THE END OF MANDATORY RETIREMENT AND FACULTY PRODUCTIVITY IN SUPERVISING GRADUATE STUDENTS During the time that I supervised it, the Office of Institutional Research was involved in a project that related to the end of mandatory retirement. Although the law ending mandatory retirement for tenured faculty members effective January 1, 1994 had been passed in 1987, Cornell, like many other institutions, had not seriously thought about what the end of the law would mean for it. When we began our research in 1997, Cornell had two years of experience living with the end of mandatory retirement. Simply inspecting the data on faculty retirement patterns we found that the vast majority of Cornell faculty were continuing to retire in advance of age 70—this suggested to us that expensive buy-out plans that paid faculty to retire prior to age 70 would be a very costly way to influence faculty retirement behavior. We also found, however, that a large share of those faculty members hitting age 70 were now staying on rather than retiring. The Office had previously developed a faculty flow model—a Markov chain model of the flows of faculty into and out of different age groups each year—and we used this model to predict what the impact of changes in the retirement behavior of those faculty still employed at age 70 would be. We found that postponement of retirement, even by this relative small share of our faculty, would substantially reduce the rate of new faculty hires at the university and would also reduce the size of the salary pool that becomes available for continuing faculty salary increases when departing senior faculty are replaced by lower paid junior faculty. This led to the appointment of a joint faculty/administrative committee, which I chaired, that developed a set of proposed changes in Cornell’s policies that were ultimately adopted by the institution. Our study and the changes made in Cornell’s policies are described in an article published in 2000.11 One thing that we did not consider in that article was how changes in the age distribution of Cornell faculty would influence graduate education at the university. In a forthcoming paper, two undergraduate students of mine obtained data on the faculty members who supervised each Cornell Ph.D. dissertation over a 7-year period from the Cornell Graduate School.12 They found that even at a major research university, many faculty members had no graduate student supervisory responsibilities during the period and also that the distribution of responsibilities was much more heavily concentrated among a relatively small number of faculty members in the social sciences than it was in the physical sciences. From the faculty data base that is maintained by our Office of Institutional Research, they were give access to each faculty member’s age, gender, field, rank, date of receipt of Ph.D. and whether the faculty member came to the university directly with tenure. The students were able to merge the Graduate School data with the data on the characteristics of faculty members to estimate, among other things, how supervision of graduate students varies over faculty members’ careers. On average, they found that the number of Ph.D. students that a faculty member supervises tends to peak at about 20 years after he or she receives the Ph.D. and that the number declines thereafter. Whether this pattern will be altered as the age distribution of our faculty changes is an open question, but it should lead the university to worry about whether an aging faculty will have an adverse affect on graduate education. COLLECTIVE BARGAINING BY GRADUATE ASSISTANTS The first graduate student union to be recognized as a collective bargaining agent was at the University of Wisconsin in 1969. By 1999, teaching assistants at 18 public Research and Doctoral Universities were covered by collective bargaining agreements and in some cases these agreements covered research assistants at the same campuses. Since the start of 1999, 16 additional major research universities and doctoral universities have recognized graduate student bargaining agents, including all the campuses of the University of California and NYU, which in 2001 became the first private university at which collective bargaining for graduate assistants takes place.13 Most universities that have been faced with graduate student unionization campaigns have vigorously opposed them, including many universities that have long had collective bargaining agreements with faculty or staff unions. The reasons for this opposition have included the belief that a system of shared governance is preferable to one of confrontation, the fear that graduate student unions might try to get involved in academic decisions that are more properly left to faculty and administration, and the fear that graduate student unions will impose financial costs on universities that will force the universities to make cutbacks in other areas and/or to increase tuition by more than they otherwise would prefer to do. Surprisingly, as of 2002, there had been no studies undertaken of the effects of graduate student unions on economic variables. Data from a data exchange conducted by a set of major universities allowed me to conduct some preliminary analyses of this question. Under the condition that I would not divulge the name of any individual institution participating in the exchange, or even the name of the data exchange, I was granted access to five years of data on the salaries, compensation, and costs of teaching and research assistants at a set of public universities. In a paper that three undergraduate students and I coauthored, we grouped these universities into three groups.14 The first consisted of 16 institutions that have never had a collective bargaining relationship with graduate assistants. The second group consisted of eight institutions that had collective bargaining agreements with their graduate assistants before 1995 and two more institutions that first began bargaining with graduate assistants in 1995 or 1996. The final group consists of seven institutions that first began bargaining with their graduate assistants during the 1999– 2001 period. For simplicity, I restrict my attention to teaching assistants in what follows, but the data for research assistants yielded very similar results. In our paper, we presented tabulations of the mean values, across institutions in each group, of a number of economic variables for five academic years, 1996–1997 through 2000–2001. The variables we looked at were 1. Average academic year teaching assistant stipend. 2. Average academic year stipend minus in-state tuition and fees that teaching assistants had to pay. 3. Average teaching assistant stipends for summer teaching. 4. Average stipends deflated by cost of living measures (either housing prices, or assistant professor salaries). In each case we compared the changes in the mean values that occurred during the period for those institutions that were either organized before the start of the period, or became organized in 1995 or 1996, to the changes in the mean values for those institutions that were never organized, or that became organized at the end of the period. On balance, we found no evidence that becoming organized during the period, or being organized before the period began led to a greater growth in academic year stipends during the period. Whether this reflects the inability of graduate student unions to win large salary increases for their members, differences in the tightness of the state budgets between the states in which institutions in which graduate students were organized are located and the tightness of budgets in states in which institutions with graduate students that are not organized are located, or a concerted effort by nonunion schools to raise stipends to try to encourage graduate students not to think about organizing, could not be determined from our analyses.15 We did find evidence, however, that graduate students who were at institutions in which bargaining was in place either at the start of the period, or began during the period, saw their required tuition payment go up by a smaller amount during the period, which suggests that graduate student union may be able to influence tuition remission decisions. Similarly, we found evidence that summer teaching stipends for graduate students increased by more at the institutions in which bargaining was in place either at the start of the period or began during the period. However, the magnitude of each of these ‘‘effects’’ was small. Hence, although these results to not take into consideration other factors that may influence theses outcomes, such as changes in state appropriations and changes in graduate tuition levels, they do suggest that graduate assistant unions have not had major impacts on the finances of public universities. Other comparisons that we reported in the paper attempted to control for differences in the change in cost of living in different areas during the period. We did so in a number of ways, including simply looking at graduate assistant stipends relative to the average salaries of assistant professors at the institution. When we did this, we found little support for the proposition that graduate student unions increase the salaries of teaching assistants relative to the salaries of assistant professors. This finding should not be too surprising—several of the graduate student contracts specify that the salary increase that their members are to receive will be equal in percentage terms to the increases granted to the faculty. Taken together our findings suggested that the impact of graduate assistant unions on economic outcomes has not been very large and the fear that graduate assistant unions will impose substantial financial costs on universities may well be overstated.16 Indeed, attracting and retaining top graduate students is an important objective of faculty at all research universities and so the faculty is often supportive of increased stipends for graduate fellows and assistants. Concern about graduate assistant unions, for the most part, is an administrative, not a faculty concern. The conditions governing my gaining access to these data limited the sophistication of the analyses that I could conduct with them. CHERI currently has a survey in the field asking graduate deans at public research and doctorate institutions to provide us with a longer historical data series on teaching assistant stipends, tuition remission policies and health insurance coverage (which was missing in the data to which I was granted access). These data will be merged with other publicly available institutional data sets and information on the date that collective bargaining coverage began, if ever, for graduate assistants at each institution. The merged data will then will be used by a doctoral student who is seeking to explain the historical pattern of the growth of collective bargaining coverage for graduate assistants and to analyze more precisely what the effects of collective bargaining coverage have been on graduate assistants’ stipends, tuition remission policies and health insurance coverage. CONCLUDING REMARKS Space does not permit me to discuss a major evaluation CHERI is conducting for the Andrew W. Mellon Foundation on the effectiveness of their Graduate Education Initiative. This initiative provided over $80 million dollars of financial support over a decade to 50 humanities and associated social science programs at ten AAU institutions to improve their doctoral programs. Data were collected annually for all students who entered both these programs and a set of control programs that were located at the ten institutions and at number of other institutions. Students were followed each year until attrition or degree completion, with information being reported on their status each year and the types of financial support they were receiving. That these data have been collected suggests that the factors that caused the cancellation of the earlier AAU/ AGS data exchange have been overcome. I mention this evaluation because I suspect that many of the offices of institutional research at universities that have departments that are either ‘‘treatment’’ or ‘‘control’’ departments in the study are unaware of both the Mellon Program and the evaluation that we are undertaking. I make this conjecture because one of the problems that institutional researchers at many institutions face is that they are not aware of all of the different data bases that are being collected and maintained on their campuses. If we are serious about using offices of institutional research to improve decision-making at our institutions, the development of a data warehouse at each institution that includes all of the institution’s databases is absolutely essential. So too is educating administrators at these institutions about the usefulness of institutional research. I take it as a personal failure that when I was a Cornell Vice President supervising our Office of Institutional Research I did not spend enough time trying to educate other key administrators about the importance of institutional research. One best unnamed senior Cornell administrator often told me that a major challenge that he faced was ‘‘managing’’ the person to whom he reported. I suspect that this is a challenge that many institutional researchers face, how do they convey to the senior administrators for whom they work the importance of what they do if these senior administrators are not ‘‘data driven’’ people themselves? Perhaps pointing out to the administrators how institutional research at other institutions has informed decisions is a useful strategy. Perhaps simply illustrating how existing cross-institutional databases have aided institutional decisions at one’s own institution is another route to follow. Indeed, as an ‘‘outsider’’ to the institutional research community, it is worth my stressing again the real benefits that multi-institutional data bases that are available to institutional researchers can provide to both the individual academic institutions at which they are employed and the broader higher education community. Two more examples, will illustrate this point. First, colleagues and I have recently used institutional level data from IPEDs and NSF to analyze what the impacts of increasing institutional expenditures for research are on institutions’ student/faculty ratios, substitution of lecturers for tenure-track faculty, undergraduate tuition levels, generation of external funds for research and annual giving.17 Second, in another paper a graduate student and I recently used institutional level panel data from IPEDs and the annual AAUP Salary Survey to trace the growing use of non tenure-track full-time and parttime faculty and to analyze the extent to which this growth could be explained by the changing relative costs of different types of faculty and by contractions in institutional budgets.18 In follow-up work we are using panel data from IPEDS and the College Entrance Examination Board’s Annual Survey of Colleges: Standard Research Compilation to analyze the impacts of changes in the proportion of faculty at an institution that are part-time and non tenure-track full-time on the institution’s 6-year graduation rate. Our analyses suggest that increased usage of both of these types of faculty members is associated with reductions in graduation rates.19 Surely both administrators and trustees at individual institutions, as well as state legislatures and governors who control state appropriations to public higher education, need to have such information in formulating their policy decisions. One of my Cornell colleagues in institutional research recently told me that the office would never have thought to conduct many of the studies that I have conducted because they are so constrained by all of the mandatory reports that they must prepare annually for their administration and for government agencies. So my final suggestion is that offices of institutional research spend some time educating social scientists on their institutions faculty, especially those who are serving on financial policies committees, about the data bases to which the offices have access and the questions that might be addressed with such data. While some might argue that I am unique and that it is rare to find a social scientist who is interested in conducting research on questions of relevance to higher education at his or her own institution, a glance at my web page should indicate to you that I had a very successful career as a labor economist before I began conducting any higher education related research.20 Indeed, my first paper on the economics of higher education, the one on optimal financial aid policies that I described at the start of this paper, grew directly out of my work on a faculty committee that was worrying about what our undergraduate financial aid policies should be. I was able to write the paper because of the willingness of Cornell’s Dean of Admission and Financial Aid and its Office of Institutional Research to grant me access to the data needed for the study. If institutional researchers and administrators think that scholars like me are hard to find, it may well be because many of them have not sufficiently drawn on the faculty at their institutions for help and encouraged them to work on research that will be both beneficial to their institution and academically useful to the faculty members’ careers.

Publication output is critical to academic staff and institutions in general. It is generally known that peer-reviewed publications are the primary unit by which faculties and educational programmes are judged. It enables academic staff to share insight, demonstrate scholarship, recognise creative thinking, and develop a reputation and expertise in a discipline. Publication output partly determines local and international recognition and respect for academic staff and institutions. Universities now also use the number of publications to an individual’s credit to measure competency (Owan & Owan, 2021). Publication output is very significant in the lives of academic staff; hence, their promotions are almost entirely dependent on it (Bassey et al., 2007). The heavy reliance on publications as the basis for research assessment has put immense pressure on academic staff to publish or perish (Glick, 2016; Lambovska & Todorova, 2021). The phrase “publish or perish” is a harsh reality for those who teach at universities (Rawat & Meena, 2014). Thus, researchers are considered productive by the number of research papers published. Many researchers define research productivity (RP) by equating it with the “quantity” of published works academics have accumulated, or what is commonly termed “publication counts” (Akbaritabar et al., 2018; Butler, 2003). Similarly, a researcher is considered productive based on the number of books or articles translated from a foreign to a local/native language, the number of written reports for consultancy work, and the number of research involvement or creative works developed (Bai, 2010; Eam, 2015). Several other scholars (e.g. Agarwal et al., 2016; Carpenter et al., 2014; Okon et al., 2022) have considered the quality aspect of academics’ RP by using bibliometric methods (such as citation counts, citation rates, h-index and others) to determine the scholarly impact of a specific author or publication. Citation index and h-index measure RP by precisely capturing the average number of times a scholar’s published works and the research works published by a university, country, region or continent are cited internationally (Kpolovie & Dorgu, 2019; Owan & Owan, 2021). The two indices demonstrate how much each faculty member contributes to the totality of human knowledge and reveal those whose research has distinctly stood out by being frequently picked up and built upon by other scholars. Today, scholars’ reputations and research excellence are determined based on the extent to which their h-index and citation counts exceed those of other scholars in the same discipline. Therefore, the more the number of scholars with an unprecedented h-index and citation counts in a university, the more prestigious the institution is publicly perceived (Kpolovie, 2014) and the higher the research income it attracts. Going by these, we can say that the productivity of academic staff is currently a “game of numbers”. It is a game of numbers because everything revolves around metrics that are not without limitations. The drawbacks to using metrics (such as citations, impact factors of journals, altmetrics and h-index) for research assessment include bias towards older work, English-language journals, discipline-specific variations, self-citation bias, lack of context, and bias towards quantity over quality. Scholars have discovered that citations and h-index can be influenced by publication venue and language biases (Cheek et al., 2006; Owan & Owan, 2021; Urlings et al., 2021). For example, researchers in fields that publish in high-impact factor journals or in English may receive more citations, making it difficult to compare researchers from different disciplines or countries accurately. Additionally, metrics do not consider the quality or relevance of the citations received (Ding et al., 2020; Dinis-Oliveira, 2019). A researcher can receive many citations for low-quality or irrelevant work, which can skew h-index, journal impact factors and other metrics. Furthermore, both citation counts and the h-index can be manipulated through self-citation, citation cartel, and other unethical practices, which can artificially inflate their scores (Loan et al., 2022; Oravec, 2019; van Bevern et al., 2020). To address the weaknesses associated with research metrics for assessment, some scholars have advocated that for a researcher to be considered productive, he must possess a solid research orientation upon starting a career in the university, have the highest terminal degree, and demonstrate early publication habits (Horta et al., 2016; Kpolovie & Onoshagbegbe, 2017; Ndege et al., 2011). He must also maintain a record of previous publication activity, communication with colleagues, subscriptions to many journals, and sufficient time allocation for research (Finkelstein in Oyeyemi et al., 2019). Other factors contributing to researchers’ productivity include affiliation to a university or research institution, assigning ample time for faculty to conduct research and using an assertive participatory management approach (Bland et al., 2005). Based on the measures of RP, preliminary assessments of the research productivity indicators of most academic staff in Nigeria tend to give a weak impression of their reputation. While it is possible to see many scholars elsewhere with h-indexes higher than 300, some scholars in Cross River State are still struggling to reach an h-index of 5. Again, most academic staff have yet to win a grant even after attaining professorial status. Some junior and senior academic staff often struggle to publish in highly rated peerreviewed journals, especially those published by well-respected publishers. Even though RP is an important requirement for research assessment, there has been a low turn-up of lecturers participating in writing and publishing research works, especially in Nigerian universities. This trend is also often observed during appraisal for promotion, where some academics are denied promotion for not meeting the minimum publication requirements expected. Again, the extent to which many scholars rely on what may be termed the “abeg put my name” syndrome in climbing their career trajectory is high (Abimbola et al., 2021; Owan & Owan, 2021). “Abeg put my name”, a phrase in pidgin English, which translates to “please add my name”, refers to a situation or system where lazy academics/individuals plead with their hardworking or productive colleagues to enlist or include them as coauthors in research papers they did not make any intellectual contribution to their creation. It is commonly associated with academics seeking to avoid perishing due to their inability or unwillingness to publish (Owan & Asuquo, 2022). Consequently, many early career academics and junior faculty members tend to be seen without the zeal to engage in active research initiatives due to their perceived reliance on other scholars for subsistence. This affects their research capacity in initiating a research idea, collecting and analysing data and preparing research reports for publication. This trend also reduces the number of scholars in universities who can carry out independent research, especially if more seasoned scholars retire from the system. Previous studies have revealed various factors affecting the RP of academics or explained why academics do not write for publication. These include the overall non-satisfaction levels of academic staff with the job (Friday & Okeke, 2020; Jameel & Ahmad, 2020; Lambovska & Todorova, 2021), socialisation of faculty staff members in a research climate (Nguyen, 2022) and university mission vis-à-vis academic research (Ghabban et al., 2019). Other factors are supervision and institutional prestige (Shen & Jiang, 2021), demographic variables (age, gender, experience, academic rank) (Anyaogu & Iyabo, 2014; Farooqi et al., 2019; Hedjazi & Behravan, 2011; Okonedo et al., 2015; Susarla et al., 2015), time allocation (Barber et al., 2021; Milem et al., 2000), teaching load (Maharjan et al., 2022; Nur-tegin et al., 2020), research competency (Prado, 2019), and interest/autonomous motivation in doing research (Masinde & Coetzee, 2021; Stupnisky et al., 2022). Furthermore, institutional factors that affect RP include poor working environment (Adetayo et al., 2023; Li & Zhang, 2022; Vuong et al., 2019), extra administrative duties, institutional support (Uwizeye et al., 2022), institutional culture and inadequate mentoring (Okon et al., 2022), among others. Although the factors mentioned above have been identified as important predictors of RP, few studies in developing nations have investigated their potency among academics. Undeniably, knowledge derived from a given context may be extended and used in another. However, care must be exercised to ensure that both contexts are culturally, economically, socially and politically similar. Given that most studies on RP are foreign to Africa, particularly Nigeria, there is a need for further studies in Africa to be conducted. This will address the current population gap in the literature and promote the applicability of research results for problem-solving. Based on this argument, the present study was conceived to determine whether mentorship and institutional support (IS) can contribute to the RP of academic staff and to determine whether collaboration and institutional culture (IC) can mediate the relationship. 1.1 Mentorship Mentoring affords the transfer of skills that protégés can apply in diverse professional circumstances. It promotes productive knowledge, clarity of goals and roles, career growth, job satisfaction and success, salary increases and promotions (Okurame & Balogun in Undiyaundeye & Basake, 2017). Previous studies on mentorship and research productivity (RP) reveal a significant positive correlation between the two variables (Abugre & Kpinpuo, 2017; Arkaifie & Owusu-Acheampong, 2019). This result implies that improvement in the mentorship practices within an institution is connected to increased productivity. However, these studies did not explain the role that other extraneous or confounding variables play in the nexus between the two variables. Many variables often co-occur, and if some are considered over others, it could skew the results, leading to misleading conclusions. The present study bridged this gap by introducing two suspected confounding variables — IC and collaboration to the relationship between mentorship and RP. Another study by Carmel and Paul (2015) indicated that mentees were positively impacted by opportunities related to career advancement, expanded thinking, scholarly confidence, facilitation of a collaborative culture, and the importance of goal setting in academia. However, the study did not consider institutional variables’ role in the established mentor–mentee relationship. This leaves us wondering whether to count on mentorship as a factor accounting for most of a researcher’s productivity. The present study addressed this gap by enlisting other variables, such as institutional culture, support and collaboration, to determine a series of interconnections among them. Research conducted by Arkaifie and Owusu-Acheampong (2019) revealed that mentoring programmes positively affected mentees’ work and personal lives. The cited study’s focus was on mentoring programmes at the institutional level. In contrast, the present study sought to determine the contribution of mentorship at the personal level among senior faculty in enriching or otherwise the productivity of mentees (who, in most cases, are junior academics). In educational psychology, Okon et al. (2022) recently found important linkages between two mentorship practices (cloning and apprenticeship) and the RP of early career researchers in South–South Nigeria. However, Chitsamatanga et al. (2018) discovered that mentoring is theoretically in universities but practically surrounded by grey areas. These grey areas include a general lack of interest and knowledge, academic perceptions of mentoring and networking in universities, and a shortage of role models. These issues promoted disintegration, inaccessibility and egoism within universities. Similarly, a study identified a problem associated with mentorship in universities by revealing that it is useful most often to the senior partner in the union (Okurame in Undiyaundeye & Basake, 2017). However, it remains unclear whether such benefit by the mentor occurs in the short or long term. It is also essential to understand when mentees will likely benefit from the relationship, as documented by Okurame. Thus, further research is still needed in this area for verification and comparative purposes. 1.2 Institutional support Institutional support (IS) refers to active organisational encouragement through policies, regulations, and monetary/non-monetary assistance offered to employees that propel them to perform their responsibilities effectively. Any organisation, including higher learning institutions, that want to earn employees’ commitment must be ready to give adequate support (Owan et al., 2022). IS in higher education can be offered through conference sponsorship, research grant provisions, publication, and technical and pedagogical support (Al- Enazi, 2016). It has been proven empirically by Salau et al. (2018) that meaningful work and growth opportunities are predictive factors for maximising productivity in institutions. However, the study was not explicitly focused on workers’ research productivity (RP), and the institutions studied were non-academic. Therefore, a study on IS and academic staff RP is necessary among academic staff populations since they are among the mainstream producers of knowledge. Research conducted by Henry et al. (2020) revealed that age cohort, the highest qualification, cluster and track emphasis are variables that significantly determine the RP of academic staff. The cited study concluded that personal, environmental and behavioural factors influence RP among academic staff. Nevertheless, none of the enlisted variables was considered concerning other confounding variables that could mediate the association. Another study concluded that even though organisational factors are significant antecedents of university academic staff RP, some of its elements (such as technological progress and possession of computer skills) were more significant antecedents than others (Hiire et al., 2020). This implies that to boost the RP of the academic staff, university managers need to place proportionate emphasis on these factors if they are to create an enabling research environment in their institutions. Some earlier studies revealed that organisational factors and funding could positively influence RP in universities and colleges (Kyaligonza et al., 2015; Musiige & Maassen, 2015). This scenario was not significantly different from what Oyekan (2014) reported in his study that resource situation factors such as physical, human and material resources have significant positive relationships with RP among academics. A recent study showed that institutional factors (availability of research funding, level of institutional networking, and the degree of research collaborations) and individual factors (personal motivation, academic qualifications, and research self-efficacy) are associated with RP in tertiary institutions in Africa (Uwizeye et al., 2022). Furthermore, Falola et al. (2020) showed that research, pedagogical and technical support are predictors of faculty responsiveness to quality research production, knowledge sharing, and administrative efficiency. This suggests that increased support to academics from their institutions might be associated with increased productivity from research engagements. Since several meaningful connections have been established in the literature on IS and RP, the factors that can mediate the link between IS and RP remain to be seen. None of the previous works has used the mediation approach to determine the roles played by confounding variables in the relationship between IS and RP. It is essential to understand all the variables contributing to a researcher’s productivity and their various roles amid other predictors. This study attempts to understand the conditions that certain IS variables are likely to contribute (significantly) or otherwise to the RP of academic staff through the mediation of collaboration and institutional culture (IC). Understanding the factors that mediate the link between IS and research productivity (RP) can provide valuable insights for academic institutions and policymakers. By identifying the specific ways in which IS affects RP, institutions can better allocate resources and create policies that support the research efforts of their faculty. Also, the use of mediation analysis in the study is a unique and important approach. This study’s use of this method can provide a more comprehensive understanding of the relationship between these variables. Collaboration is a key aspect of many research projects, and understanding how IS affects collaboration can provide insight into how institutions can support the research efforts of their faculty. Similarly, IC can shape the overall research environment of an institution and understanding how it affects RP can provide insight into how institutions can create a supportive culture for research. 1.3 Collaboration The advancement of scientific knowledge demands that researchers be equipped with the appropriate competencies, beginning from learning about the problem under study (which permits the individual to carry out original and relevant studies), up to the necessary competencies in methodologies and reporting the findings for publication. A single scientist does not possess all the necessary competencies to achieve scientific advancement in an era characterised by complex problems that transcend one discipline for effective solutions (Abramo et al., 2017). Collaborating with research teams across the affected disciplines is a good and practical strategy to overcome such problems. This is possible since collaboration would involve scientists with complementary skills and abilities. Moreover, collaboration facilitates the generation and selection of original ideas due to the synergies obtained from scientists with complementary backgrounds or even from different disciplines (Abramo et al., 2017; Rigby & Edler, 2005). Multiple authors’ involvement also permits more efficient use of time. It limits the need to resort to external advisors, for example, for thirdparty checking of research processes and outcomes (Barnett et al. cited in Abramo et al., 2017). There is a dearth of empirical studies on collaboration and research productivity (RP). However, one related study by Alaa and Ahmad (2020) indicated that funds, collaboration, ICT and job satisfaction positively impacted RP. It was further revealed that funding has the highest impact on RP. This study implies that the management of universities should pay greater attention to research funding opportunities, reward collaboration among researchers, promote ICT integration and improve job satisfaction to boost the RP of the academic staff. However, there is a lack of empirical literature on the topic. Nevertheless, the current study builds on the findings of Alaa and Ahmad (2020) by using collaboration as a mediator variable between mentorship, IS, and RP. This approach can provide a more comprehensive understanding of how collaboration affects RP and identify specific ways collaboration can boost RP. Additionally, the study can provide valuable insights for university management. For instance, Alaa and Ahmad (2020) indicated that funding has the highest impact on RP; however, this study aims to investigate how other variables, such as mentorship and IS, can also boost RP through collaboration. By understanding how these variables interact, university management can better allocate resources and create policies that support the research efforts of their faculty. 1.4 Institutional culture Several studies have investigated the organisational factors that significantly relate to the research productivity of university academics, and they have all emerged with diverse and, at times, contradictory findings (e.g. Kyaligonza, 2015; Mugimu et al., 2013; Musiige & Maassen, 2015; Oyekan, 2014). For instance, Bland et al. (2005) tested the ability of the individual, institutional and leadership variables to influence faculty research productivity using data from the University of Minnesota Medical School. Regression results revealed that faculty productivity was influenced more by institutional than personal characteristics. Hadjinicola and Soteriou (2006), on the other hand, investigated factors that promote the research productivity of researchers in US Business schools. Their findings revealed that the availability of a research centre, funding from external sources for research purposes and better library facilities are factors associated with increased RP in U.S. Business schools. In another study, Hedjazi and Behravan (2011) discovered that institutional characteristics, such as a network of communication with colleagues, facilities, corporate management, and transparent research objectives, were significant predictors of the agricultural faculty members’ research productivity. Also, Kyaligonza (2015) also revealed that institutional factors had a moderate effect on the research productivity of the academic staff in the universities studied. In another development, Ayesha et al. (2021) found that different institutional elements such as research procedure of departments, job and compensation, assets and helping materials have a low, however, positive correlation with research profitability among teaching personnel. Furthermore, Jaime (2020) established that every unit increase in holding high-performance expectations and providing intellectual stimulation could produce a 0.36 and 0.67 growth in faculty members’ research efficiency. Factors such as “nature of school leadership”, “modelling behaviour”, “providing individual support”, and “strengthening school culture” likewise contributed to faculty members’ research production but not to a significant level. Based on a literature review, there is a lack of consistency in the findings of previous studies on the relationship between institutional culture (IC) and the research productivity of academics. Some studies have found that IC significantly impacted RP, while others found that personal characteristics had a greater influence. Thus, there is a lack of consensus among previous studies. Furthermore, research on the specific role of IC as a mediator in the relationship between mentorship and RP and also between IS and RP is lacking. Additionally, a lack of studies specifically focused on public universities in Nigeria makes the current study important to fill this gap in the literature. Furthermore, studies also found that different institutional variables such as research procedures of departments, compensation, assets and useful materials have a low yet positive correlation with research productivity (Ayesha et al., 2021), highlighting the need for further research in this area. 1.6 Research questions Based on the review of related studies and the gaps identified, the current study was particularly designed to provide answers to questions such as: 1. How much do collaboration and institutional culture jointly and partially mediate the relationship between mentorship and the research productivity of academic staff? 2. How much do collaboration and institutional culture jointly and partially mediate the nexus between institutional support and the research productivity of academic staff? 1.7 Hypotheses 1. The indirect effect of mentorship on the research productivity of academic staff, with collaboration and institutional culture as joint and partial mediators, is not significantly different from zero. 2. The indirect effect of institutional support on the research productivity of academic staff, with collaboration and institutional culture as joint and partial mediators, is not significantly different from zero. 2 Methods 2.1 Research framework and design This study adopted the quantitative research method, following the correlational design. The quantitative research framework was considered due to the variables’ nature and how they were measured. The correlational research design was explicitly adopted because it was in the interest of the researchers to test for several relationships among the predictors, mediators and endogenous variables. Furthermore, “the data used to test mediation analysis are usually correlational, and such data are limited in their capacity to yield clear conclusions regarding causality” (Iacobucci, 2008; p.4). 2.2 Variables of the study We considered five specific variables in this study—mentorship, institutional support, collaboration, institutional culture and research productivity. Mentorship and institutional support are the predictors of the study; collaboration and institutional culture are the mediators; the research productivity of academic staff is the criterion variable. Following are the operational definitions of these variables. Mentorship is operationally defined as the guidance and support provided by more experienced or senior researchers to less experienced or junior researchers. This guidance and support can take various forms, such as providing advice on research projects, offering feedback on drafts of papers, introducing mentees to potential collaborators or funding sources, and providing advice on navigating the academic job market. become successful and productive researchers in their own right. Institutional support (IS) is operationally defined as the resources, policies, and practices that academic institutions provide to support the research activities of their faculty and staff. It includes variables like access to research funding, research facilities, and specialised equipment, as well as support services like research administration, library resources, and IT support. It can include academic freedom and protection for researchers, policies that promote work-life balance, and opportunities for professional development. Collaboration is defined as the process of working together with other researchers in the same or different field of studies in the same department/institution or other institutions to conduct research, analyse data, and publish findings. Collaboration can be a key driver of research productivity because it allows researchers to pool their expertise, share resources, and benefit from their collaborators’ diverse perspectives and skill sets. Institutional culture (IC) refers to the shared values, norms, and beliefs that shape the attitudes and behaviours of individuals within different academic institutions. It includes the mission and vision, values and priorities, as well as the expectations and incentives placed on faculty and staff in academic institutions. IC can profoundly impact research productivity because it can shape the attitudes and behaviours of researchers in ways that either support or hinder their ability to conduct high-quality research. For example, a positive IC that values and encourages research productivity can foster an environment in which researchers feel supported, motivated, and empowered to conduct their work effectively and vice versa. Research productivity (RP) refers to the ability of researchers to conduct highquality research and produce a significant volume of scholarly output. More specifically, RP includes publishing papers in reputable journals, presenting research at conferences, securing grants and funding, number of undergraduate and postgraduate research supervised, and current h-index and citation counts in recognised databases. Research productivity can be influenced by various factors—the availability of resources and support, the level of mentorship and guidance provided, the opportunities for collaboration, and the overall culture of the institution. 2.3 Study participants The population of this study comprised 327 academic staff from the rank of Lecturer II to Professor at the University of Cross River State (UNICROSS). This eligibility criterion was set to this range to consider the views of both seasoned and early career researchers. Besides, to reach the rank of lecturer II, staff must have earned their doctorate degrees and are fully prepared for research engagements. It is generally believed that many individuals can start research engagement after obtaining a doctoral degree. However, the actual engagement in research can vary from person to person and field to field. Hence, the decision to focus on this population was to avoid potential bias or skewing of the data obtained. The population distribution of the study is presented in Table 1. Considering the manageable number of elements in the population (see Table 1), the census approach was adopted to study the entire population. Thus, the participants of this study are 327 academic staff. 2.4 Instrument and measures The instrument used for data collection was the “Career Empowerment and Research Productivity Questionnaire (CERPQ)”. The tool was designed by the researchers and structured into six sections. Section 1 elicited respondents’ biodata such as age, gender, rank and years of work experience. Sections 2 to 6 are designed with six items each to assess mentorship practices, IS practices, collaboration, IC, and RP. Two sample items for mentorship are: My mentor provides me with useful guidance on all my research projects; My mentor helps me to identify new research opportunities in my field. Two sample items for IS are: my university provides adequate financial resources to attend international conferences; my university easily provides researchers with access to the resources they need to conduct their research (such as research facilities, equipment, or support services). Two sample items for collaboration are: I have worked with researchers from other departments within my university on different research engagements; I find it difficult working with researchers from fields that are completely different from mine. Sample items for IC include: my university encourages academic staff to take risks in pursuing new ideas in their research; my university recognises the hardworking/productive academic staff through annual awards. Two sample items for RP are: last year I published more than ten journal articles in Scopus- indexed journals; I have secured over one research grants/external funding in the past three years. All the items in Sects. 2 to 6 of the CERPQ were developed using information from a related literature review. A four-point Likert scale was used to organise all the items in Sects. 2 to 6 of the CERPQ. The response options ranged from Strongly Agree to Strongly Disagree. 2.5 Validity and reliability The researchers adopted the quantitative approach to content validity to determine how the items pooled in the questionnaire were rich, relevant, and precise items in measuring targeted constructs. A group of six independent assessors (three research experts and three educational management experts) at the University of Calabar were consulted. Experts in these two fields were used because we considered the fields to be the most closely related to the study’s variables. The experts’ ratings were used to compute the Item Content Validity Index (I-CVI) and Scale Content Validity Index (S-CVI) of the instrument (see Table 2). The researchers followed the experts’ comments on unclear and irrelevant items was subjected to a focus group discussion (FGD) involving ten senior university lecturers from the University of Calabar who were not part of the study’s targeted respondents. Qualitative discussions were held regarding the suitability of each item to the targeted domain, the adequacy of the items measuring each variable and possible omissions. Information from the FGD was used to develop the instrument’s final draft. The final draft of the instrument was trial tested on 30 lecturers at the University of Calabar (who were neither part of the study’s population nor the FGD) to determine the degree of its internal consistency. Cronbach’s Alpha approach was adopted for the reliability test, with the result of the analysis presented in Table 2. 2.6 Data collection and analysis procedures Primary data were collected for this study by administering copies of the questionnaire. Per national and institutional regulations, ethical clearance is not required for survey-based studies (Federal Ministry of Health, 2007). This is because even though the study involved human subjects, no physical, emotional or social risk was associated with participation. However, before collecting data, respondents were assured that the data solicited would be treated aggregately and used purely for academic reasons. All respondents participated voluntarily in the exercise after clear explanations of the study’s objectives were also provided. At the end of the exercise, the researchers gathered data from 303 respondents with the support of three research assistants. Twenty-four targeted respondents could not be reached for one reason, leading to an attrition rate of 7.4%. Nevertheless, the participation rate of 92.6% was considered high enough to proceed with data analysis. For data analysis, all responses were scored, considering differences in the wording of the items. Data were coded into a spreadsheet package according to the variables. According to the recommended Safe Harbour principles of studies involving human subjects, all respondents’ biodata were de-identified. Structural equation modelling and Bootstrapping techniques were employed to perform the mediation analysis to answer the research questions and test the study’s hypotheses. The analysis was aided using JASP, Hayes’ PROCESS Macro, and AMOS Graphic software. 2.7 Model specification The SEM for the mediation models of this study is specified as: where C = Collaboration, IC = Institutional culture, RP = Research Productivity, M = Mentorship, IS = Institutional support, β0 = Intercept, βM, βC, βIC, and βIS = partial contribution of mentorship, collaboration, IC, and IS while controlling for the effect of other variables εc, εIc, and εRP = error terms of the outcome variables such as collaboration, IC, and RP. 3 Results 3.1 Research question 1 Structural equation modelling (SEM) was performed to determine how much collaboration and institutional culture (IC) jointly and partially mediate the relationship between mentorship and the research productivity (RP) of academic staff in public universities. Figure 1 shows that mentorship, collaboration and IC jointly accounted for 34.6% of the total variance in the RP of academic staff, with 64.4% of the unaccounted portion of the variance due to other factors not enlisted in the model. The model also showed that 35.3% of the variance in collaboration among staff is due to mentorship. Therefore, 64.7% of the unexplained variance is attributable to other predictor variables outside the model. Furthermore, Fig. 1 shows that mentorship contributes 3.4% to the development of IC, with an unexplained variance of 96.6% due to other factors not included in the model. Relatively, mentorship positively predicted collaboration among staff (B = 0.591, β = 0.594, t = 12.842, SE = 0.046, p < 0.001) and IC (B = 0.179, β = 0.186, t = 3.283, SE = 0.055, p < 0.001) to a significant extent in the public university. This implies that a 1% increase in mentorship is associated with a 0.6 and 0.2% change in collaboration and IC, among other things being equal. However, Fig. 1 revealed that mentorship had a negative but insignificant contribution to the RP of academic staff in the public university (B = − 0.009, β = − 0.016, t = − 0.278, SE = 0.028, p > 0.05). Similarly, there was an insignificant positive contribution of IC to the RP of academic staff (B = 0.002, β = 0.003, t = 0.059, SE = 0.028, p > 0.05). Furthermore, Fig. 1 also showed a significant positive contribution of collaboration to the RP of academic staff (B = 0.339, β = 0.597, t = 10.326, SE = 0.033, p < 0.05). The result suggested that a 1% increase in collaboration among staff is associated with 0.6% improvement in their RP, assuming other things are held constant. Regarding the mediation, Fig. 1 shows that mentorship has a total effect of β = 0.339 on academic staff RP. Out of this effect, β = − 0.016 is direct and β = 0.356 is indirect—the indirect effect results from the joint mediation of IC and collaboration. Therefore, collaboration and IC jointly mediate the nexus between mentorship and the RP of academic staff at the public university. 3.2 Research question 2 Mediation SEM was performed to examine the extent to which collaboration and institutional culture (IC) jointly and partially mediate the nexus between institutional support (IS) collaboration and IC are jointly accountable for 35.5% of the overall variance in the RP of academic staff. Thus, 64.5% of the unexplained variance is attributable to other extraneous factors in the model. It is also shown that IS accounted for 3.6 and 2.5% of the variance in collaboration and IC, respectively. By implication, 96.4 and 97.5% of the unexplained variance in collaboration and IC are attributable to other factors not included in the model. Partially, Fig. 2 shows that IS significantly contributed to collaboration (B = 0.194, β = 0.190, t = 3.367, SE = 0.058, p < 0.01) and IC (B = 0.155, β = 0.157, t = 2.756, SE = 0.056, p < 0.001), respectively. This result suggests that, other things being equal, a 1% improvement in the IS provided to academic staff is tied to 0.20 and 0.16% improvements in their collaboration practices and IC, respectively. However, Fig. 2 shows that IS has a positive but insignificant contribution to the RP of academic staff (B = 0.050, β = 0.087, t = 1.826, SE = 0.028, p > 0.05). Figure 2 also shows that IC negatively and insignificantly predicted the RP of academic staff (B = − 0.004, β = − 0.007, t = − 0.140, SE = 0.027, p > 0.05). Nevertheless, Fig. 2 further proved that collaboration has a significant positive contribution to the RP of academic staff (B = 0.326, β = 0.573, t = 12.180, SE = 0.027, p < 0.01). In terms of the mediation, results showed that IS had a total effect of β = 0.195 on the RP of academic staff. However, the total effect was proven to be a product of β = 0.087 (direct effect) and β = 0.108 (indirect effect). The indirect effect of IS on the RP of academic staff is due to the mediation of IC and collaboration. Therefore, collaboration and IC mediate the nexus between IS and the RP of academic staff. 3.3 Hypothesis 1 The first hypothesis was tested at the 95% confidence interval and 0.05 alpha level. Table 3 confirmed that mentorship has a non-significant negative direct contribution to academic staff’s research productivity (RP). Using the bootstrapping technique to explain the significance of the indirect effect, we used the lower and upper bounds of the bootstrapped confidence interval. The rule is that the value of the indirect effect must fall between the lower and upper bounds (i.e. it must fall within the range of BootLLCI and BootULCI) to be greater than zero. The last part of Table 3 shows that the indirect effect (β = 0.200) of mentorship on the RP of academic staff, with collaboration and institutional culture (IC) as joint mediators, falls within the range of 0.142 and 0.268. This implies that the indirect effect of mentorship on the RP of academic staff, with collaboration and IC as mediators, is significantly different from zero. Based on this evidence, the null hypothesis was disregarded, and the alternative hypothesis was adopted. Therefore, collaboration and IC are significant joint mediators of the nexus between mentorship and the RP of academic staff. When we controlled for the effect of IC, it was discovered that collaboration partially mediated the relationship between mentorship and research productivity of academic staff to a significant extent (β = 0.200 > 0.143, but < 0.268). Therefore, the mediation of collaboration on the relationship between mentorship and the RP of academic staff after discounting for IC is significantly different from 0. On the contrary, IC did not significantly mediate the relationship between mentorship and the RP of academics when we controlled for the effect of collaboration. Therefore, the mediation of IC on the relationship between mentorship and the RP of academic staff after discounting for collaboration is not significantly different from 0. From Table 3, the following linear equations were fitted: 3.4 Hypothesis 2 For the second hypothesis, Table 4 shows that academic staff’s research productivity (RP) received a total indirect effect of β = 0.063 from the joint mediation of institutional culture (IC) and collaboration. The joint mediation of collaboration and IC was significantly different from zero. Consequently, the null hypothesis was discarded, whereas the alternative hypothesis was upheld. This indicated that although the provision of institutional support (IS) alone is not a decisive factor in boosting the research productivity of academic staff significantly (β = 0.050, SE = 0.028, t = 1.83, p > 0.05), accompanying it with effective collaboration and IC uplifts the RP of academic staff significantly. When we controlled for the effect of IC, the result showed that collaboration had a significant partial mediation on the nexus between IS and the RP of academic staff. This implies that the indirect effect of IS on the RP of academic staff, with collaboration as the mediator, is significantly different from zero. On the other hand, IC did not significantly mediate the nexus between IS and the RP of academic staff after controlling for the effect of collaboration; thus, its partial mediation is not significantly different from zero. The following linear equations were fitted using the result in Table 4: 4 Discussion This study used a structural equation mediation modelling to analyse a series relationships between mentorship, institutional support (IS), collaboration, institutional culture (IC) and research productivity (RP) among academic staff in a Nigerian public university. This study established that mentorship had a negative but non-significant contribution to the research productivity of academic staff. However, mentorship only promoted research productivity among academic staff when it was followed by IC and collaboration. One possible implication of this finding is that mentorship programmes alone may not adequately support academic staff in their research endeavours. Instead, a supportive and collaborative IC may be key to promoting research productivity among academic staff. For example, an IC that values and encourages collaboration, open communication, and sharing resources and expertise may be more conducive to research productivity than a more hierarchical or competitive culture. Therefore, universities should focus on providing mentorship and fostering an IC that promotes collaboration and support among academic staff. Another implication of this finding is that mentorship programmes in public universities may not be well-designed or implemented to support academic staff in their research endeavours effectively. For example, the mentorship may not align well with the mentees’ research interests or may not support their needs. This corroborates the finding of Chitsamatanga et al. (2018) that mentoring is important in universities theoretically, but the practical concept appears to be surrounded by grey areas. The result further agrees with Okurame, cited in Undiyaundeye and Basake (2017), who found that mentoring relationships were more helpful to the mentor than the mentees. However, the result disagrees with the evidence of other studies which established a positive relationship between mentorship and the research productivity of academic staff in public universities (Abugre & Kpinpuo, 2017; Arkaifie & Owusu-Acheampong, 2019; Carmel & Paul, 2015; Okon et al., 2022). The disagreements in the results are attributable to contextual variations and suggest the possibility of further research on mentorship and research productivity among academic staff in universities. This study established that mentorship significantly affected collaboration and IC in public universities. This is an important finding because it highlights the value of mentorship in fostering a positive and productive working environment within universities. One implication of this finding is that universities should place a greater emphasis on mentorship programmes. By providing mentorship opportunities for students and faculty, universities can promote collaboration and synergy among their members. This can lead to a more cohesive and effective institution, with better outcomes for students and faculty. Another implication of this finding is that universities should make an effort to ensure that mentorship programmes are inclusive and accessible to all members of the university community. This can be achieved through targeted recruitment efforts and mentorship opportunities for underrepresented groups such as women and minorities. This study further found that mentorship indirectly affected the research productivity of academic staff, with collaboration and IC acting as joint mediators. This finding is explainable since it suggests that providing mentorship opportunities for faculty can promote collaboration and a positive IC, which can, in turn, lead to increased research productivity. Specifically, the study found that when the effect of IC was controlled for, collaboration partially mediated the relationship between mentorship and the research productivity of academic staff. This is an important finding because it suggests that collaboration plays a significant role in the relationship between mentorship and research productivity. This finding implies that universities should prioritise collaboration as a key component of their programmes for academic staff. By promoting collaboration among mentees and mentors, universities can increase the effectiveness of mentorship for research productivity. Therefore, universities should create a positive IC that supports collaboration and mentorship. This can be achieved by promoting open communication, trust and mutual respect among academic staff. This corroborates the finding of Alaa and Ahmad (2020) that collaboration has the highest positive and significant impact on research productivity. Similarly, other scholars have also shown that collaboration synergises original ideas from scientists with complementary backgrounds or even from different disciplines (Rigby & Edler, 2005; Katz & Martin in Abramo, D’Angelo & Murgia, 2017). The study found that when controlling for the effect of collaboration, IC did not significantly mediate the relationship between mentorship and the research productivity of academic staff. This finding is critical because it suggests that IC may not play as significant a role in the relationship between mentorship and research productivity as previously thought. Therefore, universities should focus on promoting collaboration and mentorship as key factors in increasing the research productivity of academic staff rather than solely focusing on creating a positive IC. By providing mentorship opportunities and promoting collaboration, universities can increase the research productivity of academic staff without necessarily having to change the overall culture of the institution. The finding does not mean that IC is unimportant; it could mean that its impact on research productivity is less direct than collaboration and mentorship. Therefore, universities should continue to promote a positive IC, as it can positively impact academic staff’s overall well-being and satisfaction. After all, Hedjazi and Behravan (2011) found that institutional characteristics, such as a network of communication with colleagues, facilities, corporate management, and transparent research objectives, significantly predict faculty members’ research productivity. The second aspect of the study documented that IS has a positive but insignificant contribution to the research productivity of academic staff in public universities. This finding suggests that while IS is beneficial for the research productivity of academic staff, it may not be the most critical factor in determining their productivity levels. This could imply that universities should not solely focus on increasing IS to improve research productivity among academic staff. Instead, they should also consider other factors that may be more influential, such as the availability of funding, resources, and opportunities for collaboration. This aligns with Hadjinicola and Soteriou’s (2006) finding that the presence of a research centre, funding from external sources and better library facilities increased the research productivity and the quality of the articles published by the professors. This result does not support the study of Salau et al. (2018), which identified meaningful work and growth opportunities as predictive factors for maximising productivity in the sampled institutions. The joint mediation of collaboration and IC was proven to be significant in the nexus between IS and the research productivity of academic staff. The findings indicate that while IS alone may not significantly impact research productivity, the presence of both collaboration and a positive IC can serve as mediating factors, leading to increased research productivity. This suggests that universities should place a greater emphasis on fostering collaboration among their academic staff. This can be achieved through various initiatives such as funding opportunities for collaborative research projects, providing resources for interdisciplinary research, and encouraging staff to attend workshops and conferences that promote networking and collaboration. This finding supports the position of some earlier advocating for collaborative ties among scholars for complementary benefits (Abramo et al., 2017; Rigby & Edler, 2005). The findings of this study discovered, when controlling for the effect of IC, that collaboration has a significant partial mediation effect on this relationship. This implies that IS may lead to increased research productivity, but the level of collaboration among staff largely mediates this effect. This finding is not surprising because the involvement of different researchers in a project has been found to permit more efficient use of time and limits the need to resort to external advisors (Barnett et al. cited in Abramo et al., 2017). On the other hand, IC did not significantly mediate this relationship after controlling for the effect of collaboration. This suggests that while IC may play a role in shaping the research productivity of academic staff, it may not be as important as collaboration in this regard. These findings suggest that universities focus on fostering a culture of collaboration among academic staff to maximise the impact of IS on research productivity. This could be achieved through team-based research projects, interdisciplinary, transdisciplinary or multidisciplinary collaborations and regular workshops to encourage collaboration. Furthermore, institutions could recognise the importance of collaboration in driving research productivity and allocate resources accordingly. This could include funding collaborative research projects, supporting interdisciplinary teams and creating opportunities for staff to share their research findings and ideas. 4.1 Limitations and implications of the study Like every other study, this study has some limitations that must be considered when interpreting the results. First, the study was conducted in a single public university in Cross River State, Nigeria, and the results may not be generalisable to other universities or countries. Therefore, scholars in other contexts may consider replicating this study to determine the degree of consistency in results to promote generalisations. Second, the study relied on self-reported data from academic staff, which may be subject to biases and inaccuracies. Therefore, future studies may consider using qualitative or mixed methods approaches to enrich the results of this study with further information explaining the series of relationships established in the study. Thirdly, the study was conducted over a relatively short period. Further longitudinal research is needed to examine the long-term effects of mentorship, institutional support, collaboration, and institutional culture on research productivity. The study did not control for other potential factors that may influence research productivity, such as personal characteristics of the researchers (such as age, gender, rank, and experience), funding, equipment, and infrastructure, which may have an impact on the results. Despite these limitations, the study provides valuable insights into the relationships between mentorship, institutional support, collaboration, institutional culture, and research productivity of academic staff in a university setting. The results of this study can be useful for other universities seeking to enhance the research productivity of their academic staff. The study highlights the importance of considering multiple factors that may influence research productivity and suggests that institutions should prioritise fostering a supportive and collaborative environment for academic staff to maximise their research productivity. The study can also inform the development of mentorship programmes and the creation of institutional policies that support and promote research productivity. The results of this study have the potential to inform and guide the development of mentorship programmes and the creation of institutional policies in other universities and contexts. The findings that mentorship alone is insufficient to promote research productivity and that a supportive and collaborative environment is important for enhancing the impact of mentorship can serve as a starting point for universities to reflect on of collaboration and institutional culture in predicting research productivity can be used to inform the development of strategies for fostering collaboration and creating a positive institutional culture. For example, universities can use the results of this study to assess their current level of collaboration among academic staff and identify potential barriers to collaboration that need to be addressed. They can also use the results to evaluate their current institutional culture and determine if changes need to be made to promote a positive and supportive environment for academic staff. Furthermore, the results of this study can serve as a starting point for further research in other universities and contexts. This study only examined a single university in Nigeria, and future research could replicate the study in other universities and countries to test the generalizability of the results and gain a deeper understanding of the relationships between mentorship, institutional support, collaboration, institutional culture, and research productivity. 5 Conclusion This study was designed to estimate the direct and indirect contributions of mentorship and institutional support (IS) to academic staff’s research productivity (RP) at a public university in Cross River State, Nigeria. Two mediator variables—collaboration and institutional culture (IC), were introduced to determine their roles in the nexus between the predictors and the criterion variables. Several meaningful relationships among the variables of this study were uncovered. Based on these findings, it is generally concluded that mentorship negatively contributes to academic staff’s RP. However, this effect can be mitigated by a positive IC and a strong collaborative network. This conclusion suggests that mentorship alone is insufficient to promote RP and that institutions should focus on fostering a supportive and collaborative environment for academic staff to enhance the impact of mentorship. It is also concluded that institutional support has a positive but insignificant tie to the RP of academic staff. However, when collaboration and IC are considered, the effect of IS becomes significant. This highlights the importance of considering multiple factors that may influence the research productivity of academic staff rather than focusing solely on IS. Collaboration is the most important factor that directly predicts the research productivity of university academic staff. Collaboration also plays a mediating role in the relationship between mentorship, IS and RP. This conclusion implies that institutions should prioritise fostering collaboration among academic staff to maximise their RP.

The scientific literature is replete with papers which refer to research, scientific, or institutional integrity. Bonn and Pinxten (2019) recently published a paper appropriately entitled, “A Decade of Empirical Research on Research Integrity: What have we not looked at,?”. They found that 986 papers were published on the subject from 2005 to 2015. Perhaps the plethora of papers and the ambiguity of their findings can be attributed, in part, to the fact that, integrity, has two definitions: (a) the accuracy, completeness and consistency of data (b) the adherence to a code of moral values To which definition of integrity did the authors of these papers refer? Is an institution responsible for the, accuracy, completeness, and consistency of data from a given experiment or study? How does one evaluate an institution’s, code of moral values? These criteria are readily applied to the work of an individual or group of scientists however, how does one apply these standards to an institution? The Office of Research Integrity (2007, 2) has defined research integrity as, “The honest and verifiable methods in proposing, performing, and evaluating research; reporting research results with particular adherence to rules, regulations guidelines and commonly accepted professional codes or norms.”. The Concordat to Support Research Integrity, developed by several research organizations in the UK, also provide criteria for measuring institutional integrity in research. It states, “An institution (shall have made) a commitment to: 1. The ongoing development of a culture that supports and nurtures research integrity; 2. The presence of a commitment to mechanisms to provide assurance that when things go wrong appropriate investigations and actions are forthcoming. 3. Assurance that the institution has,” the people, systems and procedures to meet the evolving demands of providing assurance of the integrity of UK research (Concordat to Support Research Integrity 2012).” How does one determine an institution’s adherence to such standards? Martinson and colleagues have provided one instrument that indirectly measures institutional integrity, A Survey of Organizational Research Climate (Martinson, Thrush, and Crain 2013). Institution and individual perceptions by the administrators, faculty and students of their university and department policies. Utilizing this measure, they determined, not surprisingly, that more positive individual perceptions of the research climate in one’s university or department were associated with higher likelihood of desirable and lower likelihood of undesirable research practices. Senior faculty found the climate to be more favorable than their junior colleagues (Crain, Martinson, and Thrush 2013). In a study of Dutch institutions, Haven et al. (2020) had similar findings. What role, if any, by its action or inaction, was played by an institution where misconduct has occurred? Had institutional policies and/or failure of their implementation played a role? Had an administrator or senior faculty member failed to act responsibly? If so, what has been the subsequent response of the institution’s leaders to this failure? The answers to these questions serve as a measure of, institutional integrity. If the problems of flawed institutional policy, ineffective senior faculty members or administrators have not been adequately addressed, the stage remains set for future acts of research misconduct On rare occasions, institution have been penalized when it was determined that they contributed to an act of research misconduct. Flawed HIV research at Iowa State University led to a criminal conviction of the principal investigator who was sentenced to incarceration; the University agreed to repay 496,000 USD and forego 1.4 million USD in grants (Bernstein 2015). A Federal research misconduct suit against Duke University was settled with a 112.5 million dollar financial penalty (Kaplan 2019). However, such severe institutional and individual penalties are rare. To make a determination of institutional integrity we suggest the use of what might be referred to as a, research post mortem exam or perhaps, more precisely, a post hoc research misconduct inquiry to be carried out by the institution where an act of research misconduct has occurred. Physicians utilize the post mortem examination to more definitively identify the contributing factors to a death. With that knowledge in hand, they ask, what, if anything, might have been done which would have been more effective? A post hoc research misconduct inquiry, of the institution’s role carried out by senior institutional authorities, could play a similar role and lead to changes in an organization whose policies and/or administrators have failed to sustain the requisite standards. Current policy calls for the presence of misconduct to be established and the responsible individual(s) to be identified by an institutional investigation organized with the help of its Research Integrity Officer (RIO). Does the institution in question have a trained RIO? If not, has someone subsequently been identified and trained? When an institutional investigation has determined that an offense has occurred, a detailed Report is submitted for review to the Office of Research Investigation (ORI) of the National Institute of Health or to the Office of the Inspector General (OIG) at the National Science Foundation. There, final determinations are made regarding the nature and presence of research misconduct and the appropriate sanctions to be invoked. However, these institutional reports can be rejected as inadequate.(Gunsalus, Marcus, and Oransky 2019). Such inadequate reports may have been avoided had the institution a trained RIO to organize the inquiry. The presence or absence of a RIO is one indicator of, institutional integrity. Trainees supported by NIH make up the largest portion of those found to be guilty of research misconduct (Kornfeld 2019). Therefore, the post hoc research misconduct inquiry should ask, were the institution’s mentors in any way, passively or actively, responsible in the trainee’s actions? Did they provide adequate time for training, supervision, and just listening? For example, one study found that 62% of mentors had not established standards and 73% had not reviewed raw data (Wright, Titus, and Cornelison 2008) A study of Dutch scientists also found insufficient supervision to be a contributor to research misconduct (Haven et al. 2020). If a trainee were found to have personal problems or emotional distress at the prospect of a failed experiment, did the mentor provide support and/or make a referral for professional counseling? A study conducted by the Wellcome Trust in the UK found that 53% had sought, or had wanted to seek, professional help for depression or anxiety (Wellcome Trust 2020). Does the institution make such help available? If not available at that time, has the institution since established such resources? If a faculty member were found wanting in the role of mentor, what action was taken? Discontinuation as a mentor? Or, if the individual is to continue in that role, what training has been provided to enhance his/her capacity to do so effectively? Will adequate time be made available in the future for mentors to effectively play that role ? Where NIH/ORI and NSF/OIG determine that the institution’s post hoc self examination has been inadequate, the relevant body should initiate its own

TECHNOLOGY TRANSFER EXPANDS UNIVERSITIES’ capacity to act more like firms and less like research and educational institutions (Slaughter & Rhoades, 2004). Consequently, university officers who make decisions about institutional income streams from intellectual property must balance conflicting goals, such as the desire to increase external revenues vs. academic integrity and human subjects’ protection, creating greater potential for institutional conflict of interest (ICOI). In May 2009, in recognition of the risk to research integrity posed by ICOI, the Public Health Service (PHS) issued a call for public comment on how to define and address I C OI (U.S. Department of Health and Human Services, 2009). To better understand ICOI within research universities, we locate university ICOI within the broader literature on conflict of interest (COI) and institutional review boards (IRBs), introduce possible dimensions of ICOI, and then analyze the ICOI policies of the 60 U.S. Association of American University (AAU) universities. The AAU is an association of 60 leading public and private U.S. research institutions.1 Institutional conflict of interest is a growing concern for research universities. Between 1995 and 2005, over 1.5 million patents were granted to U.S. universities (National Science Foundation, 2008). As university ownership of intellectual property grew, so too did licensing income and the numbers of spin-off companies based on academic inventions. In 2005, according to the Association of University Technology Managers (AUTM), 24 universities reported earning more than $10 million from licensing income. A handful earned over $100 million. The total number of reported new start-up companies in 2005 was 404 (Blumenstyk, 2007). Conflict of interest has been studied primarily at the level of individual researchers (Boyd, Lipton, & Bero, 2004; Bekelman, Li, & Gross, 2003; Cho et al., 2000; Van McCrary et al., 2000). Studies of COI at the institutional level have generally focused on Institutional Review Boards (IRBs) (Campbell et al., 2006; Campbell et al., 2007; Bartlett, 2008; Wolf et al., 2008; Reeser et al., 2008) and occasionally on ICOI that IRBs are asked to address, for example, when IRBs are asked to review potential ICOI by university officials who are considering making an investment in intellectual property based on faculty research that involves human subjects (Weissman et al., 2008). COI policies focus on a variety of conflicts that may confront faculty, often concentrating heavily on ownership of intellectual property, protection of human subjects, and faculty interaction with industry. As Cho et al. (2000) point out in a study of COI policies directed largely toward faculty and research staff, including postdoctoral researchers and graduate students, increased involvement with business “has created opportunities for conflicts of interest for university faculty members because academic-industry partnerships can offer direct financial rewards to individual faculty members in the form of consulting fees, royalties, and equity in companies while simultaneously funding these faculty members’ research” (p. 2204). The concern is that the promise of financial gain will consciously or unconsciously influence faculty members’ research in terms of quality, outcomes, and dissemination (Krimsky, 2003). The potential for financial COI adds another dimension to concern about human subjects’ protection in that researchers running clinical trials may have direct or indirect interest in the financial outcome if they hold patents on treatments, processes, or medical devices. COI and ICOI ICOI is somewhat different than COI, and there has been little research on ICOI. ICOI refers to situations in which research, teaching, or service are compromised because external financial or business relationships held at the institutional level may bring financial gain to units or the institution in form of increased revenues, whether payments or donations, or when external financial relationships have the potential to influence decision making regarding these activities. Of course, individuals may benefit as well when institutions license their patents or invest in their products. As the University of Kansas ICOI policy states: An institutional conflict may develop when the institution (such as a department, center or college, the applicable Research Foundation, or the University) stands to benefit financially from the outcome of research ongoing at the University to support a license or a research agreement. A Research Foundation, and/or units at the University, along with inventors, may receive future financial rewards by way of royalties or other fees if the product or service is commercially successful. Therefore, they have a financial interest in ensuring the success of the product (University of Kansas, 2007). Administrators at the rank of department chair and above are the focus of ICOI policies. Although faculty are often the originators of intellectual property, administrators, specifically university officers, are usually the institutional actors in positions that allow them to influence or appear to influence research investments in ways that maximize individual or institutional gain, even though they are not directly involved in conducting research. All AAU research universities now have some combination of human subjects’ research policies, IRBs that administer those policies, individual COI policies, and COI committees. Many also have ICOI policies. These ICOI policies have evolved as universities have entered into greater numbers and varieties of university-industry partnerships, including industry-sponsored research, technology licensing based on patents, and investment in startup companies. These various boards, committees and policies overlap to some degree and differ from university to university in their scope. For example, some universities have a separate COI committee to which IRBs refer apparent COIs and ICOIs to determine whether they are permissible and if so, how they are to be managed; other universities require their IRBs to identify and manage COIs and ICOIs, though some (e.g., Weissman et al., 2008) have expressed concern about whether this responsibility should be added to IRBs’ already considerable workload and whether IRB members have sufficient training to be able to evaluate or manage COIs and ICOIs. At present, there are no clear federal guidelines specifying roles, so for example, IRB members surveyed on their opinions regarding investigators’ relationships with industry did not agree on whether investigator-industry relationships should be considered when reviewing protocols, and among those who believed that such relationships should be disclosed to research participants, only about half of those members reported that this is done (Weissman et al., 2008). As COI scholars have pointed out, the various and overlapping policies and committees may contribute to the lack of knowledge and uncertainty about processes that many respondents to CIO surveys exhibit (Campbell et al., 2007; Weissman et al., 2008). With regard to medical schools, 25 of the 125 medical schools inform their IRBs of potential ICOI if human subjects are involved (Ehringhaus et al., 2008). POTENTIAL DIMENSIONS OF ICOI Potential ICOI has several possible dimensions: (1) university as firm in which institutional officers are economic actors who manage university technology transfer activities to maximize commercial interest; (2) “sand and gravel” in which an officer acts to maximize their own self-interest; (3) exchanges in which institutional officers engage in broad, unspecified quid pro quo arrangements with industry. (1) The university as firm conflicts are the most recent type of ICOI and arise primarily when universities are involved in commercialization activity. This may range from sponsored research to licensing intellectual property or holding equity positions in spin-offs based on faculty research (Krimsky, 2003;Washburn, 2005; Bozeman & Hirsch, 2005; Kavanaugh, 2009). In these transactions, institutional administrators act as corporate rather than academic managers. When university officials participate in commercialization, they become economic actors, seeking to maximize revenues. This economic goal may conflict with research integrity or human subjects’ protection requirements, hence the need for ICOI policies. Equity agreements are particularly contentious because financial markets react to information to which academic institutions have privileged access. There is the potential for researchers and university officials to manipulate share prices by withholding information or grandstanding. This potential ICOI is intensified because of the substantial value of the investment if the company is successful. Whether university officials are licensing an intellectual property or taking equity in a corporation based on faculty research or initiating an initial public offering (IPO), the economic goal is to maximize revenue and that goal may be in conflict with the goals of research integrity or human subjects protection, hence the need for ICOI policies. University as firm conflicts may also compromise academic values of openness and sharing of scientific information. Faculty and/or graduate students may be asked to withhold or delay publication while universities acquire patent rights to intellectual property (Slaughter, Archerd, & Campbell, 2004). When universities own intellectual property, a desire for profits may lead them to restrict access to the knowledge they now own. For example, Harvard’s oncomouse, licensed to DuPont, was a proprietary cancer research tool, the use of which had to be negotiated between DuPont and the Public Health Service in order to protect DuPont’s investment before it became widely accessible to researchers (U.S. Department of Health and Human Services, 1999). Moreover, universities often lobby state government on legislation that restricts public access to universityindustry agreements on the grounds that it would benefit competitors. However, this secrecy impedes public oversight (Slaughter & Rhoades, 2004). The Jesse Gelsinger case illustrates a university as firm conflict of interest. Gelsinger was an 18-year-old man recruited to participate in a gene therapy clinical trial at the University of Pennsylvania, despite contraindications that should have excluded his participation. Gelsinger died while participating in the trial. The lead investigator, James Wilson, a professor, along with William Kelley, former medical school dean and then Chief Executive of the University’s Health System, held three patents on aspects of the experimental gene therapy. These patents were licensed to Genovo, Inc., a company founded by Wilson that also provided research funding for his university laboratory. In return for the funding, Genovo had the right to develop commercial products from the research. Wilson and the University of Pennsylvania also held stock in Genovo. As Washburn (2005) points out, “[B]oth Wilson and the university stood to profit if the study was successful.” And, despite Gelsinger’s death, both Wilson and the University of Pennsylvania profited. Genovo was sold to a larger company, making Wilson’s stock options worth $13.5 million, and the university’s equity share worth $1.4 million. When faced with litigation by the Gelsinger family after Jesse’s death, the University of Pennsylvania reached an undisclosed, out-of-court settlement. In this case, the institution was financially involved through its funding of research, the conduct of the clinical trial, and its investment in Genovo. While research funding and oversight of clinical trials are covered by COI policies directed primarily toward faculty and researchers, the equity investment in the firm based on faculty intellectual property created a situation in which the university could gain financially if the trials were positive. The potential of financial gain creates a conflict with objective inquiry and research integrity. University officers are in positions that give them oversight over this range of conflicting activities, which are governed by ICOI policies. The Gelsinger case exemplifies the interplay of roles that ideal ICOI policies should attempt to target. Without James Wilson, the faculty member, there would have been no intellectual property that resulted in the creation of the firm in which the university took an equity position. William Kelley, the former medical school dean, who supervised Wilson, was listed as an inventor on three patents and was involved as both a researcher and administrator, but was apparently acting as an individual rather than an institutional representative with regard to patenting and oversight for the experimental therapy. Technology transfer officials as well as responsible senior officials acted as representatives of the university as firm when the University of Pennsylvania took an equity position and later when a settlement with the Gelsinger family was negotiated. It is sometimes very difficult to separate faculty roles from administrative and managerial roles, and individual roles from institutional roles. Even the best ICOI policies reflect these difficulties. (2) Sand and gravel.2 University officials have always faced the potential for sand and gravel, or procurement conflicts. A classic example is a vice-president for finance who owns a business and uses their influence to steer a lucrative university contract to their business without or despite competitive bidding. The office puts individuals in a position to seek improper financial gain for themselves or their families and friends. These conflicts arise because of the individual’s position as a university officer. General state COI laws for public universities cover this type of conflict. State statutes that deal with non-profit organizations cover university officials at private universities. 3. Quid pro quo exchanges occur when a unit within the university, or the university as a whole, has an exchange relationship with industry brokered by university managers. Through this relationship, industry provides various forms of support to the university in return for association with the university’s name or university use of its products. Sometimes these relationships are contractual and clearly spelled out. At other times, they are ill-defined or tacit exchanges. Quid pro quo exchanges frequently involve university departments or units. Campbell et al. (2006) found in their survey of medical schools and teaching hospitals, two-thirds of the departments as administrative units had quid quo pro relationships with industry. These relationships involved research equipment, grants of unrestricted funds, support for research seminars, support for training and residency for students, support for department-administered continuing medical education (CME), discretionary funds for food and beverages, support for professional meetings, and subscriptions to professional journals. A conflict arises because these relationships may commit the department to certain activities when other companies and products might better meet the therapeutic, educational, and even financial goals of the university. Quid pro quo exchanges also occur at the level of the university. These exchanges are often unspecified and sometimes tacit. A donor holds out the promise of funds and university officials act in ways they think will secure them. The ICOI issue is similar to that at the department or college level: the university official(s) may act in ways that secure the gift rather than in ways that serve human subjects or the public interest. At the extreme, as in the Olivieri case at the University of Toronto, university official(s) may undermine the integrity of the research process. In this case, the Canadian pharmaceutical corporation Apotex had promised the University of Toronto a multi-million dollar gift, and the university initially did not support Nancy Oliveri, a professor and researcher, when she attempted to report negative findings about Deferipron, an Apotex product (Thompson, Baird, & Downie, 2001). As institutions became involved in increasing numbers of academic-industry relations and these become more complex, the need to develop ICOI policies that address all three dimensions of ICOI became apparent. Associations representing research universities began to address the topic of ICOI policies after 2000. In 2008, the Association of American Universities (AAU) and the American Association of Medical Colleges (AAMC) issued a template for ICOI policies that focuses specifically on human subjects’ research (Association of American Medical Colleges Task Force on Financial Conflicts of Interest in Clinical Research, 2002; AAMCAAU, 2008). There has been commentary but little empirical study of COI or ICOI policies (McKinney & Korn, 2005). The exception is a recent survey of deans of medical schools (Ehringhaus et al., 2008). The survey focused on two key dimensions of ICOI as identified in the AAMC-AAU policy statements: “those (dimensions) created by financial interests held by the institutions and those potentially created by the financial interests of institutional officials.” The study found that adoption of ICOI policies by U.S. medical schools was “far from complete on both dimensions” (p. 669). Rather than examining ICOI policies, the study queried the deans on pre-formulated questions addressing the two dimensions. Because the deans were not queried about institutional policies per se, they could respond only according to their knowledge of policies or practices, and they were not asked to distinguish between the two. This paper takes a different approach, analyzing the ICOI policies at AAU institutions on multiple dimensions to better understand the rules universities have devised to regulate ICOI. Although this type of analysis does not provide information about how specific cases at institutions are handled, it systematically reveals the procedures and tools available to administrators monitoring ICOI, and allows for comparison and assessment.3 Method The sample for this study consisted of ICOI policies that we collected from the websites of the 60 U.S. AAU universities. Data were collected between September 2005 and September 2007. Because of the AAU universities’ extensive involvement in commercialization, they are an appropriate group for the study of ICOI. In 2006, the 53 AAU universities in the Association of University Technology Managers (AUTM) survey constituted 23 of the top 30 institutions in terms of licensing revenue, capturing 84.5% of that revenue, and accounted for 90% of the startups in the same group (AUTM, 2006). 6 S. Slaughter, M. Feldman, S. Thomas Although paper copies of ICOI policies identified by institutional officials may have been the preferred data source, we had difficulty obtaining such copies. Our initial efforts to identify appropriate officials and policies were unsuccessful. When we contacted officials at public universities, they often evinced lack of clarity about which ICOI policy was authoritative, while officials at private universities were reluctant to reveal any information, sometimes stating they were not required to do so because of their private status. When officials cooperated, they most frequently referred us to websites. It is certainly possible that we did not locate the right officials, but time constraints led us to decide that a search of websites would be more efficient than more telephone calls. A web search is a justifiable way to locate ICOI policies because the Web is very likely the first place a faculty member, researcher, or institutional officer would turn to find relevant ICOI policies.4 As the AMMCAAU (2008) states, “Institutional COI policies should be publicized on campus and made available to the public” (p. 19). We understand that all institutions likely have policies that at the very least apply to officials in a position to benefit financially from their university posts, whether these are in state statutes governing non-profits or part of the state administrative code. Moreover, we reasoned that ICOI policies that were difficult to locate were unlikely to have a powerful impact on the persons or positions to which they were directed. We developed a search strategy that we applied to each institutional website. The following search terms were used: COI or COI policy or policies; faculty COI policy or policies; ICOI policy or policies, trustee/regent ICOI policy or policies, or, if another term, such as Board of Governors, designated the oversight body that was used. Although ICOI policies were the target, we used the broader search terms to ensure that ICOI policies located at various institutional levels and sites were captured. For example, general COI policies occasionally contained material about ICOI, while faculty-directed COI policies sometimes contained language that applied to institutional officers. Similarly, trustee/regent COI policies had some language that covered institutional officers. In addition to a general search at each university, we also searched the following pages and/or websites for policies addressing ICOI: research; faculty handbook; human resources; medical school; researchmedical school; and regents/trustees. If the university was part of a system, we searched for ICOI policies on the system’s web pages. As we searched the Web for ICOI policies at AAU universities, we discovered they had numerous and at times multiple policies that covered COI generally. However, our primary concern was locating and analyzing policies that regulated institutional officers involved in financial activity that could lead to potential ICOI. When reviewing a “community” policy that addressed all members of the university, some institutional officers at the level of department head or above had to be specified by office if the policy was included in our sample. Similarly, sections of faculty policies were included if “faculty and officers” were named, although we extracted data only about the “officers.” Given that more than one policy at any single university often dealt with ICOI, we typically extracted data from the policy that governed other policies. For example, a number of university medical schools had ICOI policies, as did the universities in which they were located. Within universities, various colleges and even departments occasionally had ICOI policies. When a university policy indicated that it provided the guiding framework within which other policies were articulated, that was the policy from which we extracted data. At public universities in a system (e.g., the University of California system and the University of Colorado system), the system level often provided the framework within which component parts had to operate, and then we extracted data from the system level. We excluded policies that dealt with conflict of commitment, which we take to be “conflict over competing faculty responsibilities, e.g., whether faculty allocate more time to their traditional academic duties or their industrial sponsor” (Campbell & Slaughter, 1999, p. 311), or the many other conflicts that university policies currently address. We converted web pages to Adobe PDFs to standardize page length. Of the 34 public institutions surveyed, 26 (76%) had publicly accessible ICOI policies on their websites. Of the 26 private universities, 18 (69%) had such policies. Two of the public institutions had 2 sets of ICOI policies, applying to different groups of officers, as did 2 of the private institutions. For example, at several institutions, one set of policies covered technology transfer officers, while another set covered “senior officers,” such as vice presidents and the president. In other words, of the 60 AAU universities, 44 (73%) had publicly accessible online ICOI policies; the 44 institutions had 48 ICOI policies. A data abstraction instrument was developed that covered topics related to ICOI, including five major categories: (1) general descriptors, (2) prohibitions, (3) disclosure processes, (4) review and management processes, and (5) other characteristics. There were 19 subcategories within these five major categories (see Table 1). Results EFFECTIVE DATE We located effective dates for all but 2 policies. Eleven universities established ICOI policies before 1995, 3 between 1995 and 1999, and the remainder (30) after 2000. The dates for the policies may not be reliable, in that the dates given may be the date of last revision rather than the date of initiation. Regardless of whether dates are for initiation or revision, the majority of activity for private and public universities occurred after 2000. LENGTH Universities’ policies that addressed ICOI in community or faculty and officers policies ranged from 1 or 2 sentences to 20 pages. Private universities’ policies were generally brief, the majority less than 1 page. In contrast, public universities’ policies were longer, averaging 11 pages. That public universities, which frequently have well-developed bureaucratic structures and are sometimes embedded in state systems, have longer policies than private is not surprising. COVERED GROUPS The ICOI policies covered the following groups—the university community as a whole (community); faculty and officers; trustees and officers—or were separate policies that dealt specifically with ICOI and referred specifically to officers or managers, who could range from the level of department head to the president but were most often “senior officers.” Of the 44 institutions with policies, 18 had community policies, 5 had policies that covered faculty and officers, 9 had policies that covered trustees and officers, and 12 had separate policies. Of the 8 separate public policies, 3 were directed to the university research foundations, on which the senior officers (vice presidents and above, as well as senior technology transfer officers) sat, while only 1 of the private policies was aimed at a university foundation. While private universities generally imposed fewer prohibitions, public and private universities were more similar than dissimilar with regard to the majority of prohibitions. Both addressed sand and gravel issues such as prohibiting officials or managers from engaging in real estate transactions involving the institution. Both generally specified delay periods before officers and managers who terminated employment with the institution could again do business with the university. However, there were a few notable differences in what public vs. private institutions prohibited (see Table 2). Public universities had many more prohibitions that dealt with purchasing, the use of positions to obtain political office, the acceptance of gifts and misuse of insider information than did private institutions. For the most part, these sections of the public policies read like sand and gravel policies directed toward public officials—in this case, using public office to secure private gain. “Outside employment” clauses were somewhat different. For the most part, they read like sand and gravel policies, allowing no outside employment in order to prevent managers from double dipping. However, outside employment clauses were occasionally directed toward officers acting as managers of university as firm, prohibiting directorship and management positions that might conflict with enhancement of the university’s portfolio. Another clear distinction between public and private universities centered on ICOI policy clauses about officers having financial interests in businesses transacting with universities. A greater percentage of private universities (6 universities, or 100%) had prohibitions than did public (9 universities, or 60%). Universities had prohibitions about administrators or officers who represented the university as firm having financial interests in faculty research (see Table 2). Generally, these prohibitions centered on what officers and administrators representing the university as firm were prohibited from doing when seeking to expand revenue streams from the university. For example, they specified the following provisions: • Technology transfer officers were prohibited from engaging in licensing agreements when the institution was a partner in a venture fund. • Universities were prohibited from investing in firms in which faculty were managers. • Universities were prohibited from being the lead investor or syndicating agent if faculty research was involved. • Universities were prohibited from holding more than 10% of equity in a faculty startup company. • University officers were prohibited from being company officers or from holding equity prior to an IPO. • Officers and managers were prohibited from investments in companies if they supervised faculty members who started such companies. The universities also had a number of prohibitions about investment decisions that impinged on faculty research activity. These prohibitions focused on universities not making decisions about technology investments or upstream investments that committed the universities financially to technology lines if such commitments could in any way potentially jeopardize the opportunity for faculty exploring alternative research leading to competing technologies. Four universities had prohibitions against clinical trials if the institutions held equity positions in a product or treatment. Three private universities (50%) had such a provision compared to one public university (6%). The provisions were aimed at preventing potential ICOI or the appearance of potential ICOI when institutional officers and managers acting for the university as firm stood to have the institution benefit from successful trials. These policies specifically addressed the conduct of clinical trials for products from university equity startups. Some policies prohibited any clinical trials under these circumstances while others stipulated procedures at different stages of the clinical trial. Although a number of AAU universities’ ICOI policies had clearly articulated prohibitions, they also had specific language in their policies that created exceptions to the prohibitions. For example, policies that prohibited universities from taking equity positions in products or treatments when university researchers were running clinical trials nonetheless usually had provisions that outlined circumstances in which the universities could take such positions despite the prohibition. For example, exceptions could be made if the particular institution was the lone facility with the scientific expertise to conduct a study, or if the public interest was best served by running the trials at the institution. Explicit language about exceptions may make recognition of exceptions more likely. Exceptions to prohibitions and disclosure processes will be more fully discussed below. DISCLOSURE Of the 44 institutions with ICOI policies, 37 (84%) required that officers and administrators disclose financial interests that might create conflicts. In the policies requiring disclosure, individual officers or administrators initiated the disclosure process, generally by filling in ICOI forms annually, or if their situation changed and a potential new conflict emerged, updating the form. Only two policies, both at public institutions, called for institutional committees or staff to proactively monitor for potential problem areas likely to trigger disclosures. In these policies, institutional committees or COI staff were required to scan university investments— particularly equity—corporate gifts, donations and research funding, technology transfer activity, and research involving human subjects to determine whether universities and/or faculty were investing based on university-owned/faculty-discovered intellectual property, or if corporations were funding research in areas in which the institution or faculty had commercial activity or were conducting research on human subjects. The committees or COI staff then alerted faculty and officers about potential dangers, in case they had missed them. However, there was no discussion of followthrough after the alert. Generally, the policies required the individual, whether faculty, officer, or administrator, to disclose not only their financial interests, but also the financial interests of immediate family members (parents, spouse, and children) in any business that did business with the university, whether that business had contracts with the university for goods and services, or whether the business involved technology transfer. These policies were sand and gravel, directed at preventing officers and managers from using family members with whom they shared finances to acquire inappropriate interests in university businesses. Disclosure entailed revealing financial interests, covering a wide range of activity, ranging from investments to positions on external boards (see Table 3). Sixteen public and 9 private universities specified financial interests to be disclosed by institutional officials and managers in the ICOI policies. The 16 public universities had a wide range of monetary activity that triggered disclosure— from holding more than $500 or 5% interest in a business or 3% in an equity company to $20,000 or 5% interest in an entity. The private policies had a similar range, but were more nuanced, specifying different types of investments and activities, with the range starting slightly higher than the public’s, with the exception of 1% of securities. The same number of public and private universities asked officers and managers to disclose partnerships, managerial positions, and trusteeships. Neither consulting nor outside activity were clear triggers for disclosures. A single public university required that the institution qua institution disclose if it had invested more than $100,000 or had a 5% interest in a company. Although only 26 of the universities (59%) specified what financial interests and activities triggered disclosure by officers and mangers in ICOI policies, disclosure may be dealt with elsewhere in university documents. Even if officers and managers were not specifically mentioned in the ICOI policy, they nonetheless may be expected to fill in disclosure forms, which are usually separate from the policies. These forms, often available on the Web, covered more employees than those named in the policies, were very detailed, and may have had many more disclosure triggers than those in ICOI policies. Disclosures were most often made to and reviewed by senior officers or university boards of trustees or regents and were quite similar for public and private universities (see Table 4). Nineteen (40%) of the policies required disclosure to be made directly to senior officers or members of boards of trustees/regents. If the 5 policies in which disclosures were made to legal counsel, university officers, and/or boards of trustees and regents, and the 2 in which disclosure was made to senior officers and/or system senior officials are combined with the disclosures made directly to senior officers or members of boards of trustees/regents, then 54% of the policies called for disclosure to and review by high-ranking officials, or high-ranking officials and senior legal counsel. Very few made disclosures to an ICOI committee. After disclosure and review, decisions about management plans were sometimes made if conflicts were identified. The officials or committees that made the plans were quite similar to those who received and reviewed the disclosures (see Table 4), except that there were fewer in the “other” category and significantly more (8, or 17%) that employed an ICOI committee that worked with senior officials and/or boards of trustees and regents. More private universities (5, or 25%) used this process than public ones (3, or 10%). Final authority over the disclosure process was highly concentrated. “Other” officials and committees dropped out and, if senior legal counsel, senior university officers, and boards of trustees/regents are combined with senior officers and board members, 34 policies (72%) conferred authority at the highest levels of the university and/or its oversight board (see Table 4). The disclosure process makes clear that ICOI policies focused on senior-level officers, who disclosed either to those immediately above them (dean to provost, provost to president, president to trustees) or to each other (technology transfer officers to vice-presidents; vice-presidents to provost or president, depending on the lines of authority; provost or chancellor to presidents; and presidents to senior system officers). The focus on administrators and senior officers underscores the individual and institutional potential for ICOI at high levels. When the individual is the target, the policies focus on sand and gravel potential for ICOI, in which an individual can use their position to enhance business opportunities inappropriately. When the university as firm is the target, the policies are aimed at managers and officers with the authority to act as agents for the institution, seeking to prevent them from enhancing revenues at the expense of the university as a community, at the expense of human subjects, or at the expense of the public good, as well as to prevent inappropriate commercial activity. MANAGEMENT PLANS Although over three-fourths of the ICOI policies called for disclosure on the part of administrators and senior officers, approximately half of these policies (49%) did not discuss how management plans would be developed to manage conflicts that were identified. More private university policies (11, or 55%) lacked plan specifics than public ones (11, or 39%; see Table 5), but otherwise the management tools were quite similar. Recusal was the most commonly named management tool. A greater percentage of private universities (16, or 80%) relied on recusal than did public ones (21, or 75%). In most policies, recusal meant that the managers or officers with financial interests in decisions had to absent themselves from the decision-making process. However, in some instances, managers and officers were able to present their views on financial decisions to the decision makers prior to the decision. In other words, managers and officers could be heard, although they could not participate in voting. Other management tools appeared in 7 or fewer (14% or less) of the policies. Creating a firewall between officers and managers who handled investments and those who managed technology transfer or endowments was a tool for managing potential conflict, as were using arms-length organizations or freezing or sequestering institutional investment. If faculty, administrators, or officers were in a situation with a high potential for ICOI, such as investment in a start-up engaged in clinical trials, divesture, vacating management positions, and resiting trials were tools that could be used to minimize conflict. Disclosure of the scope of individuals’ and institutions’ financial interest in the products, treatments, or processes to the human subjects involved in clinical trials was also a means of managing ICOI. Several private universities saw independent external review committees as a way to handle conflicts, and one private university called for a review of adverse advents. In sum, the majority of universities called for disclosure but did not specify management plans or tools. Approximately half the policies did not discuss how plans to manage identified ICOI should be developed. Although the ICOI policies in aggregate provide an array of tools for managing conflicts, most universities deployed only one to three in their policies. EXCEPTIONS Eleven (55%) of the 20 private university policies had language in their institutional policies that provided for exceptions to prohibitions and disclosure processes, as did 11 (39%) of the 28 public university policies. A number of policies had multiple exceptions. The exceptions ranged from allowing acceptance of gifts in certain circumstances to allowing university investment that breached firewalls. Regardless of prohibitions, decisions about how to manage ICOI were made on a case-bycase basis. SANCTIONS AND COMPLIANCE WITH STATE LAW Private university policies did not mention sanctions for violation of ICOI policies explicitly, although two policies mentioned complying with state law with regard to ICOI. In contrast, 11 public institutions (42%) mentioned sanctions in their policies and 20 (77%) indicated that they would comply with state law in regard to ICOI. The sanctions ranged from administrative sanctions through civil and criminal ones. For the most part, the sanctions were not elaborated. Generally, state laws were not specified, but occasionally the name and number of the law were provided. Discussion We posited three possible categories of ICOI policies— sand and gravel, university as firm, and quid pro quo— prior to our analysis. After conducting our analysis with regard to the degree of coverage of conflict situations, we revised our categories to the following four policies: minimal policies; sand and gravel policies; modestly elaborated university as firm policies; and elaborated university as firm policies. Minimal policies were very short, usually only several sentences, focused on individual administrators rather than the institution, and had no more than one management item, such as to whom disclosures are made or what is disclosed. Of the universities with policies, 7 private universities (39%) had minimal policies, as did 5 public universities (19%). Universities with sand and gravel policies had several prohibitions and/or management tools, but focused on managers and officers as actors who may violate ICOI by maximizing individual gains. Five private universities (28%) had sand and gravel policies, as did 9 public universities (35%). Modestly elaborated university as firm policies regulated the activities of managers and officers acting not only for themselves but also as institutional executives seeking to maximize university commercial activity. Although these policies recognized that managers and officers made decisions that committed the university as an institution to strategic investment decisions with the potential for ICOI, they did not have well-specified tools for managing such conflicts. Two private universities (11%) and 7 public universities (27%) had such policies. Elaborated university as firm policies regulated the activities of managers and officers acting for the university as firm through a variety of specified management tools. Three private universities (17%) had such policies, as did 5 public institutions (19%). The 8 elaborated policies were all separate policies, although not all 12 separate policies were elaborated. Two of the elaborated public policies dealt with university foundations that managed technology transfer on which senior officers of the institutions sat as managers. Quid pro quo policies did not figure in AAU ICOI policies. Taken together, the procedures and management tools articulated in the elaborated policies brought together some of the building blocks that might provide a comprehensive approach for regulating ICOI for the university as firm. (Most of the prohibitions [Table 2], and tools for managing disclosed ICOI [Table 5] were drawn from the elaborated policies, and may be reviewed for details.) Among the building blocks are the following: • Disclosing to equity review committees, ICOI committees, external review committees, and to the public, professional associations, scientific journals, and human subjects; • Options other than individual initiation of disclosure; • Close regulation of officers and managers when investments were made in faculty intellectual property; • Management tools that went beyond recusal, including creating “arms-length” distance between managers engaged in activities such as technology transfer, fund-raising, investment and other financial decisions through firewalls, vacating, divesting, resiting, sequestering, freezing institutional involvement, creating independent external review committees when human subjects participated in clinical trials, and reviewing adverse events. EMPLOYMENT OF STUDENTS A single institution, one with elaborated policies, addressed the work practices and employment of graduate students when discussing management of ICOI. This public institution took the position that when the research foundation had an equity interest, it must monitor the situation with a management plan that attended to resource allocation, employment practices, and graduate student assignments. Management plans were required to be placed in individual personnel files and maintained for three years. The goal of the plan was to ensure that the graduate student was treated as a student with an array of learning opportunities rather than as a worker in an enterprise from which the university would benefit. The plan tried to ascertain whether graduate students were fairly compensated, they worked reasonable hours, and they had assignments that fostered their education. (For faculty concerns about the problems that arise when graduate students are employed on projects in which faculty and university have financial interests, see Slaughter et al., 2002). MATERIAL TRANSFER AGREEMENTS Although the elaborated policies cover many of the current commercialization activities in which universities engage, there is little mention of material transfer agreements (MTA), which govern the transfer of tangible research materials when the recipient intends to use the material for research purposes. The most frequently transferred materials are biological, such as reagents, cell lines, plasmids, and vectors, but MTAs may also be used for other types of materials, such as software or chemical compounds. The potential of financial gain has created a formal review process of each request. Each agreement has provisions to govern confidentiality, publication delay, and control of intellectual property (Mirowski, 2008). MTAs are generally handled by university legal counsel and technology transfer offices, and are infrequently incorporated in ICOI policies. However, transferring materials between organizations, which may be competitors in both a scientific and financial sense, creates a potential for conflict. ICOI MANAGEMENT TOOLS Although the elaborated plans had a number of tools for managing ICOI, monitoring of the management plans was not widely discussed—an internal audit of a random sample of plans was mentioned in only a single instance. Sanctions were not elaborated, although they were mentioned in half of the elaborated plans, mainly at public institutions. None of the policies addressed continuing education, despite the rapidly changing commercial environment in which managers and officers of research universities operated. Only 8 of the AAU universities had elaborated ICOI policies that used a range of management tools, and none of these used the full panoply. As Table 6 indicates, while a majority of the AAU universities with ICOI policies required that officers and managers disclose activities with the potential for ICOI, less than half specifically prohibited officers and managers from engaging in activities likely to lead to ICOI and less than half had management plans that offered tools to deal with identified ICOI. A considerable number of institutions specified exceptions to prohibitions and a relatively small number of universities articulated sanctions. As Cho et al. (2000) noted with regard to faculty, the policies we analyzed with regard to ICOI varied widely among institutions and many lacked specificity with regard to the types of relationships with industry that were allowed or prohibited. MULTIPLE ROLES In a number of the ICOI policies, the lines between faculty, individual managerial, institutional, and trustee conflicts of interest were blurred. While these various categories of employees overlap, given that faculty are sometimes institutional managers and occasionally managers of firms, ICOI policies often do little to clarify the ways in which conflicting statuses may exacerbate conflicts. COMPLEXITY OF CASES AND FALLIBILITY OF POLICIES Although ICOI policies are important, they may not be sufficient to handle all of the issues raised by ICOI at research universities. As Bozeman and Hirsch (2005) pointed out in their study of the Johns Hopkins lead paint case, “the greater the complexity of the social phenomenon, all else being equal, the greater the fallibility of the rules” (p. 82). They argued that the Hopkins’ IRB had technical scientific sophistication but was not able to deal with all of the ethical implications of different levels of abatement of lead paint in houses inhabited by young children in poor neighborhoods in Baltimore, which led to IRB approval of a research design in which some young participants’ lead blood levels went up due to the type of abatement assigned. ICOI decisions are certainly as complex as the IRB decisions made in Johns Hopkins lead paint case, but may be more difficult to monitor because they are usually not made in as public a forum as an IRB. INTEGRITY OF SENIOR OFFICERS Because of the non-routine nature of commercial activity at research universities, all ICOI is managed on a case-by-case basis. That means institutional policy is only as strong as the various committees or senior officers responsible for developing and monitoring management plans to handle conflicts. ICOI committees, when universities have them, are usually somewhat different from IRB committees, but IRB committees nonetheless offer insight into conflict management. As Campbell et al. (2007) recently found, 15.1% of surveyed IRB members at research universities reported that at least one protocol came before the IRB sponsored by a company with which they had a relationship or a competitor company, and that 19.4% of the conflicted board members voted on the protocol regardless. ICOI committees responsible for monitoring conflicts may find similar difficulties as research universities become more involved in commercialization. ICOI policies can be no better than the committees or officers and managers that handle them. Creating an organization culture attuned to ICOI issues may be as important as developing ICOI policies. However, there are two problems with creating an organizational culture attuned to ICOI. First, ICOI policies, as this study reveals, are directed primarily toward high-ranking officers and managers. IRBs and faculty are generally not the focus of ICOI policies. These highranking officers and managers are not accustomed to close regulation, and indeed would undoubtedly claim they required discretion and authority to make independent investment decisions for successful returns. Even when elaborated university as firm policies provided tools for management plans, for the most part, senior officers and managers disclosed only to each other, and/or to boards of trustees or regents, raising the question of who is governing the governors. Second, when ICOI policies address officers and managers acting as agents for the university as firm, institutions may not want elaborated rules controlling investment in intellectual property that may constrain the ability to expand external revenues. Best Practices It is possible to approach best practices from a reasoned position, as well as drawing on some of the management tools in the elaborated ICOI policies. Conflict of interest is difficult to avoid or eliminate, but it may be managed. TRANSPARENCY The increasingly dense, complex relationships between research universities and industry make the case for transparency. Information about university investments in commercial firms and financial interdependencies between private research sponsors, donors, and those who stand to benefit from university activity decreases the likelihood and/or appearance of ICOI. University trustees, regents, and advisory board members are often selected because they work with or represent corporations that have interests in industrial sectors salient to universities, e.g., biomedicine, and already have ongoing complex relationships with the university or because they are in a financial position to be donors. In other words, the stewards of the universities stand a good chance of being interested parties when it comes to research and institutional investment. The best practice would be to provide publicly available information on potential conflict situations so that it is possible to monitor them. Transparency as a practice fits well with increased public concern about financial regulation following the 2009 crisis of the global financial system. An annual listing of company investments in institutional intellectual property, whether patents, copyrights, or trademarks, that provided the names of products, processes, and services, and intellectual property details, such as licensing terms, would be a solid step towards transparency. Such a practice would allow individuals, units within universities, officials, and senior management as well as the university community as a whole to assess quickly potential ICOI if they pursue commercial activity as individuals employed by a university or if they act as agents for the university as firm. Human subjects, the public, undergraduate and graduate students, as well as any other interested parties could also access and assess the commercial endeavors of research universities. Such a practice would address the Association of American Medical Colleges–Association of American Universities (2008) concern that “[T]he existence (or appearance) of such [ICOI] conflicts can lead to actual bias, or suspicion about possible bias, in the review or conduct of research at the institution. If they are not evaluated or managed, they may result in choices or actions that are incongruent with the missions, obligations, or values of the university” (p. 1). Transparency could reduce the potential for ICOI without burdensome oversight. Public posting of universities’ investments would be a relatively efficient and effective bureaucratic process. Individual faculty, managers, and officers’ disclosures could be quickly and easily checked against public postings; recursively, the public postings would alert individual faculty and officers to areas of concern, causing them to examine carefully their interests in these areas. When faculty and managers, whether acting as individuals or acting as agents of the university as firm have interests in these areas, they would automatically be referred to an independent, external ICOI committee (see below). Universities already collect the information required for public posting, so existing institutional procedures and offices would be utilized. The only extra steps required would be consolidation and posting of the information, as well as checking it against disclosures. Research universities may be uncomfortable with public scrutiny of institutional commercial activity and may make the case that posting such information regularly and publicly might damage university-industry relations. If research universities do not accept public posting of commercial investment and endeavor, then the next best practice may be to shift some of the burden for reporting ICOI from individuals to the institution. Two of the elaborated ICOI policies called for institutional committees or staff to monitor proactively for potential problem areas by scanning university investments, corporate gifts, donations, and research funding, as well as technology transfer activity, with an emphasis on institutional equity investments and research involving human subjects. As with public postings, when faculty and managers, whether acting as individuals or acting as agents of the university as firm have interests in these areas, they would automatically be referred to an independent, external ICOI committee. SEPARATION OF FUNCTIONS AND DUTIES The best policies call for the separation of officers and managers who handle investments from those who manage technology transfer or endowments. Several of the elaborated ICOI policies deal with separation through external management of investments. In other words, an entity independent of the university handles investment portfolios so that the university staff neither makes nor influences individual security transactions. This type of separation typically addresses sand and gravel ICOI rather than university as firm conflicts. Sand and gravel policies focus on the individual officer who might use their position for self-dealing. In contrast, university as firm policies address the group of officers or managers making business decisions about intellectual property and how it may be used to increase external revenues for the university. The difference is similar to that between stockbrokers who direct investors toward businesses from which they stand to gain and venture capitalists representing their firm who take risks with firm funds for high returns; the firm, in this case, the university rather than the individual, is the direct beneficiary. Separation is further complicated in that a university officer may simultaneously have both individual or sand and gravel and institutional or university as firm conflicts. The potential for conflict for individual investment managers is eliminated by a firewall between the investment manager and persons who handle technology transfer activities or human subjects, because the investment manager, separated from the persons who deal with technology property or human subjects, cannot take advantage of any special or insider knowledge about university intellectual property. However, investment management by individuals occurs after investments have been made. What is at issue and is often the source of conflict for officers and managers acting as agents for the university as firm are collective decisions about institutional intellectual property that risk institutional funds for future income. Because the future of the investment is uncertain and the process of realizing the return often lengthy and complicated, with multiple decision points, the likelihood and nature of conflict is unclear. For example, university officers acting as agents of the university as firm have very likely already patented and taken equity in a drug discovered by a faculty member before decisions are made about running clinical trials. Milestones may have to be reached, more university funds may need to be committed, and decisions about more research funding made before the process is complete, with each decision point raising the possibility of ICOI, especially if ICOI is considered as involving not only human subjects but research integrity and university values. For example, the university may decide not to run clinical trials itself, but may continue to accept and commit funds to research even though successful commercialization is increasingly unlikely, using funds that could have been used more productively in other areas, from research investment for other commercial purposes to research investment for social purposes that may not realize investment returns but that would fulfill the university mission to serve the public good. Income from managerial decisions made for the university as firm may not enter the university’s portfolio, where protection from ICOI is ensured by a firewall, for years, but in the interim, university managers may face many potential conflict situations. Although all of the elaborated policies call for firewalls, most of them simply assert separation of functions and duties and do not detail how this will occur. Nor does the separation of functions and duties address the question of who supervises the supervisors. Separation of functions usually occurs below the level of president/chancellor, who are at most institutions ex officio members of the boards of trustees/regents. These senior officials and fiduciary officers act together on financial decisions and are not accountable to any outside party. Yet about 25 percent of president/chancellors sit on corporate boards, and some sit on multiple boards (Goldschmidt & Finkelstein, 2001); the majority of private- sector trustees are corporate CEOs and directors of other corporate boards, many of whom have interests in universities’ research areas (Slaughter, Feldman, & Thomas, forthcoming). As Ehringhaus et al. (2008) note in their survey of medical school deans, “Of the [35] institutions reporting the use of separate foundations for investment management purposes, licensing/technology transfer purposes, or both, 29 (56%) report that senior institutional officials, (i.e., president/chancellor, provost/vice chancellors/deans) serve on the governing boards of such separate entities and 18 (35%) report that midlevel officials serve on such boards” (p. 669). The purpose of the separate foundation may be to create a firewall, but in many cases officials and managers who have may have strong financial interests in the investment, licensing, or technology transfer decisions to be made are on the wrong side of the firewall. The potential for ICOI at this level is both sand and gravel and university as firm. An official or trustee could act individually on eminent investment decisions by universities to enhance individual worth, or officials and trustees could act collectively to invest in research with the potential to increase external revenues in areas which the corporations they represent have or could have interests. TRIGGER POINTS Best practice policies are clear about the triggers for ICOI. These include: university ownership of intellectual property or investment in areas in which the institution is running clinical trials or otherwise involving human subjects; universities taking equity positions in faculty/institutional companies; universities allowing financial or other interests to influence which technologies are selected for commercialization, who receives licenses, or the structure of licenses; universities accepting corporate sponsorship or gifts for research in areas in which the institution has made an investment or in which the institution partners with the corporation. One university suggested that if these triggers are present, an external ICOI committee composed of persons sophisticated in financial matters with an affinity for the institution, but whose financial interests are not dependent on it, form an independent committee to ensure that any research undertaken in a trigger area have no greater risk to research integrity or human subjects than comparable studies. This external committee would work with the board of trustees, the IRB, and appropriate institutional officials to provide oversight of research projects with identified ICOI. COVERED INSTITUTIONS For the most part, the best practices described here are appropriate for research-intensive universities. These institutions already gather much of the information discussed in the Best Practices section (above) and have institutional offices and procedures that already address COI or are skilled at constituting internal and external committees to address specific issues such as ICOI. Nonresearch intensive universities that do not have technology licensing offices, or hold large intellectual-property portfolios would probably be able to best address ICOI through public posting (discussed earlier), a task that would not be burdensome until such universities had large intellectual portfolios—at which point they might reconsider their procedures and policies. Research Agenda This project studied only ICOI policies, not how ICOI is handled within universities. While there are a number of studies of ICOI cases, for the most part, these focus on very public ICOI failures (Thompson, Baird, & Downie, 2001; Krimsky, 2003; Washburn, 2005; Bozeman & Hirsch, 2005). Several possibilities for research prior to public mishaps are possible. For the long term, universities could agree to embed researchers in institutions who could study how university managers and officers make ICOI decisions over time through participant-observation supplemented with interviews. For the shorter term, researchers could study NIH listings of adverse events to see if and how they are related to the ICOI policies at the institutions at which they occur. Another path would be an analysis of legal cases in which ICOI was involved, since cases are often fact-rich and court decisions shape ICOI policies, although potential in this area is somewhat limited, given that many universities settle these issues out of court, leaving no public record (Kavanaugh, 2009). And of course, surveys of faculty, administrators, and officers with the potential to be involved in ICIO are very useful. Educational Implications Senior managers and officers of universities would very likely benefit from education about their legal, ethical, and fiduciary responsibility when they act for the university as firm. University retreats might be a venue for such educational activity. A range of experts could be brought in to explore the ICOI situations that senior managers and officers may encounter, and to work through various solutions. Special emphasis should be given to the differences between sand and gravel and university as firm ICOI. The issues surrounding quid pro quo exchanges, which are not yet part of ICOI policies, should be considered. Attention should be given to the way ICOI intersects research integrity and university values, rather than concentrating only on human subjects’ research. Continuing education modules should be developed and required for mid-level managers, given the changing regulatory and legal environment pertaining to ICOI. As with much medical and legal continuing education, these modules could be offered via the Internet, through tapes, at conferences, or retreats. However, given the complexity of the issues and that ICOI is still in a nascent stage, education would probably be most successful in face-to-face forums. Care should be taken that the ethics and values of science and of the university are embodied in ICOI considerations; legal considerations alone are not sufficient to safeguard the integrity of scientific research with regard to conflicts of interest.

Scientific creativity is a key driver for scientific and technological progress, and also a precondition for advances in other societal domains. Yet, our knowledge and understanding of how research organizations and institutional environments, and changes in both, impinge upon capabilities of research groups to conduct creative research is fragmented. The complex relationships between productivity, social stratification, reward structures, and organizational context in scientific researchwere frequently studied until the mid- 1970s within the institutional paradigm of the sociology of science (see, for example, Shepard, 1956; Meltzer, 1956; Merton, 1957; Meltzer and Salter, 1962; Stein, 1962; Pelz, 1964; Crane, 1965; Pelz and Andrews, 1966; Cole and Cole, 1967;Reskin, 1977; Zuckermann, 1977; Andrews, 1979; Long and McGinnis, 1981). Since then science studies have been dominated by a social-constructivist paradigm that focuses on the micro-conditions of knowledge production in laboratory settings and epistemic cultures (Latour and Woolgar, 1979; Knorr-Cetina, 1981, 1999; Knorr-Cetina and Mulkay, 1983). At the same time, the study of creativity has become popular in psychology, although organizational and institutional questions play only a marginal, if any role (Dunbar, 1997; Amabile, 1996; Sternberg, 2003; Simonton, 1999, 2004). It was only recently that new attempts were undertaken to re-establish an organizational and institutional perspective in the study of scientific accomplishments. For example, Hollingsworth (2000, 2002) and Hage (2006) have published on organizational structures that foster breakthrough research. Hemlin et al. (2004) have explored various institutional factors that are associated with what they call “creative knowledge environments”. Yet, in their book on serendipity in science, Merton and Barber (2004) conclude that the institutional analysis of discoveries in science is still in its infancy. Many important questions remain about what creative scientific accomplishments are, how we can identify them, in which organizations they occurmost often, and whichinstitutional factors are influential in shaping cutting-edge research environments. The desire to knowmore about the factors that contribute to scientific creativity is given further impetus by the substantial changes seen over the last few decades in the institutional and organizational conditions under which scientific research is conducted (Jansen, 2007; Laudel, 2006; Schimank, 2005; Etzkowitz, 2003; Owen-Smith, 2003; Langfeldt, 2001; Bourke and Butler, 1999). Public research funding is nowincreasingly allocated through competitive processes, rather than long-terminstitutional block-grants; increased research collaboration is encouraged through a variety of measures including through organized research centers, networks, centers of excellence, and interdisciplinary teams, to address diverse challenges of complexity, convergence, knowledge exchange, scale, scope, and internationalization in contemporary science; and evaluation systems for research performance are increasingly implemented as a supplement to peer review(Münch, 2008; Thèves et al., 2007; Lepori et al., 2007; Corley et al., 2006; Shapira and Kuhlmann, 2003; Chompalov et al., 2002; van Leeuwen and Tijssen, 2000; Henkel, 1999). In the context of heightened competitive pressures to foster science-driven business development and the rise of new global locations for research (especially China), research policymakers in developed economies hope that adjustments to research organizations and broader institutional environments for scientific research will not only promote more efficiency but also boost scientific excellence and creativity (Blau, 2005). There is an increasing need for recommendations about the design of science policy to support highly creative researchers and their groups. This paper explores factors which influence the ways in which researchgroups conduct theirwork. Besides features of the research group itself, such as size and career stage of group leaders, our main analytical and empirical focus is on organizational variables and the institutional environment in which these groups operate, such as leadership, funding structures, or competitive pressures. Our study is built on a longitudinal multi-method research design based on survey, interview, archive and bibliometric data, and uses both quantitative and qualitative researchmethods including as network and regression techniques, and in-depth interviews and case studies (Heinze et al., 2007; Heinze and Bauer, 2007). We identified creative research accomplishments in two broad fields of science, analyzed why certain research groups are more creative than others, and investigated which factors in theirwork environmentwere influential for their accomplishments. We begin the paper by reviewing contributions to the literature on scientific creativity and by highlighting selected key issues important for further research (Section 2). Second, we introduce our methodology (Section 3). Third, we discuss in more detail the results from our case studies of highly creative research accomplishments, focusing particularly on findings related to organizational and institutional influences on scientific creativity including work group factors, such as size of research groups or communication patterns, and organizational features, such as sponsorship or disciplinary variety (Section 4). Then, we discuss our findings in the light of previous results, and we demonstrate howour findings improve our understanding of creative knowledge environments (Section 5). Finally, we consider the implications for research management and research policy (Section 6). 2. Literature review: definitions, approaches and findings on scientific creativity The importance of creativity in numerous areas of society has resulted in studies of creativity from diverse fields, including management and business (Sutton, 2002), arts (Maritain, 1977; Berka et al., 2003), politics (Otten, 2001; Nagel, 2002), and urban and regional development (Florida, 2002). However, there is a convergence in characterizing creativity as encompassing capabilities to do things that are new and useful (see Ochse, 1990 and Amabile, 1996, for a summary of definitions). In the world of science, creativity is similarly defined in terms of knowledge and capabilities that are new, original, surprising, and useful (Hollingsworth, 2004; Simonton, 2004). As in other fields, standards and norms are established in science against which claims for innovative contributions are assessed, although science, more than other fields, has evolved procedures, disciplines, and institutions to accredit new knowledge (Whitley, 2000). In making judgments about scientific creativity, scientific peers use criteria such as plausibility, validity, and originality. There arewellrecognized tensions here, since criteria of plausibility and validity tend to encourage conformity, while originality draws upon and encourages dissent. The history of science is replete with examples of path-breaking research achievements thatwere initially rejected by the scientific establishment because they challenged existing paradigms (Kuhn, 1962; Polanyi, 1969; Hessenbruch, 2004). In other cases,work thatwas initially proclaimed publicly to be highly creative was found, following more considered scientific review, to be flawed (Lewenstein, 1992). In short, the scientific community must be persuaded that novel and unexpected contributions have value, and claims that research is highly creative need validation over time and by other scientists. There are varied approaches to examining and empirically measuring creativity. These include examining creative individuals, the products or outcomes of creative work, creative processes and creative knowledge environments (Stumpf, 1995; Hemlin et al., 2004). At the individual level, there has been much discussion – not necessarily with consensus – about the relationship between intelligence and creativity (Mansfield and Busse, 1981; Sternberg, 2003). There has also been a focus on the behavioral traits of creative individuals, including their level of curiosity, risk tolerance, motivation, and willingness to overcome failure, leading to arguments that creative people typically tolerate higher levels of contradiction, ambiguity, and uncertainty in their work (Sternberg et al., 1997; Weinert, 2000). Still, such individual characteristics are neither easily measured nor uniformly correlated with creative accomplishments, leading others to concentrate on tangible scientific publication outcomes and citations to identify highly creative researchers. A prominent attempt to assess scientific creativity through outcomes is publication and citation analysis within an evolutionaryprobability theoretical frame (Simonton, 1999, 2004). Simonton argues that scientists who are highly productive in publishing papers encounter a greater likelihood that one or more of their papers will come to the attention of other scientists, be cited, and recognized as creative. In other words, the more contributions to knowledge that a scientist produces, the higher his or her chances are that one of these contributions resonates well in the scientific community. This approach is not without criticism because, for example, some highly creative scientists publish only a few papers, while citation counts typically consider only journal publications and not books or other contributions, such as new scientific instrumentation. Another outcome approach is based on studying prestigious prize winners in science (Zuckermann, 1977; Hollingsworth, 2002). Of course, such prizes are highly selective – and there are surely more creative research accomplishments than Nobel committees can recognize. Hollingsworth (2002) addresses this problem by obtaining access to short-listed Nobel Prize nominees until the 1940s. Creative processes, including the selection of problems, methods, partners and knowledge sources, have been another area of inquiry. Rather than focusing on innate individual traits, work on creative processes has highlighted the opportunity structures in collaboration networks that facilitate the generation and diffusion of novel ideas. Proponents of network brokerage argue that people who are placed at the intersection of heterogeneous social groups have an increased likelihood of drawing upon multiple knowledge sources, leading to the generation of new ideas. For example, managers who occupy brokerage positions are more often than others the source of good ideas (Burt, 2004; Rodan and Galunic, 2004). In contrast, proponents of cohesive collaborative networks argue for the benefits of trust, shared risk taking and easy mobilization in facilitating information and knowledge transfer. According to these studies, individuals with cohesive social ties are more likely to be involved in innovations (Uzzi and Spiro, 2005; Obstfeld, 2005). In reference to this ongoing debate, Fleming et al. (2007) argue that although brokering inventors are more likely to generate new ideas, the brokered network structure itself is less suited to diffusing these ideas. Therefore, network structures that enhance the generation of novel ideas may inherently diminish the likelihood of their diffusion Considerations of the research environment add a series of further elements – including organizational and institutional features – to the examination of scientific creativity. Research environments influence opportunities for research collaboration and multidisciplinarity, which may in turn affect processes of knowledge discovery. But once we contemplate the role of organizational and institutional aspects, a range of other factors may come into play in stimulating creative research situations, including autonomy for researchers, adequate facilities and funding, development of complementary disciplines and fields, staff selection, management structures, and leadership. Less conducive factors include: insufficient basic funding, limited time for research, bureaucratic management, narrow range of disciplinary expertise, and excessive evaluation and accountability pressures (Hemlin et al., 2004, pp. 16–17, 195–196). Perhaps the first comprehensive effort to empirically examine modern research environments was a study of 17 research facilities in the United States across various types of fields and laboratories by Pelz and Andrews (1966).1 Another major effort to analyze research organizations in a comparative fashion was the UNESCO study by Andrews (1979) on more than 1200 research groups across six countries.2 Again other studies have investigated university departments and their role in setting research goals and influencing scientists’ productivity (Long and McGinnis, 1981; Baird, 1986; Allison and Long, 1990), More recently, Hollingsworth (2002) examined a large number of research breakthroughs in the biomedical sciences across 128 research organizations in the United States, of which 28 had two or more major discoveries in the first half of the twentieth century.3 There are also historical accounts of exceptionally successful institutes in the biomedical sciences, such as Rockefeller University, California Institute of Technology (Hollingsworth, 2000, 2004) or Institut Pasteur (Hage, 2006). These studies differ in methodology, for example, with respect to the identification of research groups, by key outcome variable (productivity, recognition or research breakthroughs), and by level of analysis (group, department, institute, or project). Significantly, although some are published recently, they all examine research organizations and institutional environments from earlier periods. For example, Hollingsworth (2002) and Hage (2006) study breakthroughs in the first half of the 20th century; Pelz and Andrews (1966) report on research organizations of the 1950s; and Andrews (1979) capture the situation before 1975. Research organizations and institutional environments have changed extensively in the last three decades following periods of post-Second World War expansion (Windolf, 1997), 1970s stabilization (Ziman, 1994), and more recent restructuring (as noted in Section 1 of this paper). While our emphasis is on recent research organizations and institutional environments, it is insightful to consider the findings from studies focusing on earlier scientific generations. In particular, this literature raises three important themes still relevant today, namely: specialization, communication, and research autonomy; group size and departmental effects; and 1 Pelz and Andrews analyzed industrial, government and university labs which spanned the following R&D fields: pharmaceuticals, glass and ceramics, electronics, electrical equipment, weapons guidance, animal diseases, commercial uses of agricultural products, basic research in several physical sciences, biological and social sciences (Pelz and Andrews, 1966, p. 2). 2 Andrews and colleagues studied within the fields of mathematics, astronomy, physics, chemistry, life sciences, earth and space sciences, agricultural sciences, medical sciences, technological sciences, and social sciences the following types of research organizations: academic organizations, academies, cooperative organizations, productive enterprises, and private institutions (Andrews, 1979, pp. 17–52). 3 Hollingsworth (2002) examined universities, medical centers, free standing research institutes, and industrial research laboratories in the biomedical sciences. resources, recruitment and leadership, as discussed in the following sections.4 2.1. Specialization, communication, and research autonomy Pelz and Andrews (1966, pp. 22–27, 35) find that scientists are most productive when they both interact vigorously with and involve their colleagues in setting up their research goals. Research productivity is correlated with high frequency of intraorganizational communication. Hollingsworth (2000, 2004) argues that research breakthroughs are typical for research organizations where scientists communicate across disciplinary and thematic borders, and where research leaders provide strategies for integrating scientific diversity with rigorous standards of scientific excellence. For example, because the Rockefeller University was organized around laboratories rather than scientific disciplines and fields, it had a greater capacity to adapt quickly to research strategies and to allow effective communication across cognitive boundaries (Hollingsworth, 2004, pp. 34–35). Further strategies for intellectual integration within the boundaries of an organization are mobility of researchers and teamwork between departments (Hage, 2006). However, the way in which the individual and organizational levels interpenetrate is somewhat contested in the literature. Hollingsworth (2000) emphasizes scientific excellence and depth of domain-specific knowledge at the individual level in combination with intellectual integration at the organizational level. In contrast, Pelz and Andrews show that high-performance scientists are often not in agreement with their organization in terms of research agenda and strategy. The authors argue that “a laboratory remains vigorous when it encourages a certain tension between what the members want, and what they think the organization wants” (Pelz and Andrews, 1966, p. 139).5 Such tension, however, is only bearable if scientists share the same motivation for their work: “It seemed helpful if sets of close colleagues shared a common enthusiasm for similar kinds of problems and preferred social relations” (Pelz and Andrews, 1966, p. 146). Pelz and Andrews found thematic breadth to be most effective when combined with freedom in goal setting and research strategy,where as coordinated research settings were better suited to more specialized researchers (Pelz and Andrews, 1966, pp. 29–31, 158–173). 2.2. Group size and departmental effects There is considerable evidence in the literature that research performance initially tends to rise as group size increases, but that above a certain group size threshold, this effect tails off or becomes negative, i.e. either no increase or even a decrease in performance (for a review, see von Tunzelmann et al., 2003). For example, analyses based on the large dataset of Andrews (1979) show that above a threshold of 4–6 team members, per capita performance decreases markedly, particularly in academic natural science groups. Also, in order for groups larger than 5–7 scientists to reach the performance levels of smaller groups, both coherent research programs and group leaders with strong time commitments to research activities are needed. Although the curvilinear relationship between group size and performance is evident both for quantity and quality of research across various countries and fields, quality seems to be affected more negatively from large group size than per capita research quantity (Andrews, 1979, pp. 55–94, 192–222). In addition, the department level has been found influential. Long and McGinnis (1981) and Allison and Long (1990) find for the fields of physics, chemistry, mathematics and biology that scientists with growing departmental prestige tend to show an increase in both the number of publications and the number of citations. Again other studies find that university departments with clear research goals show higher productivity levels than those without such goals (Baird, 1986), and that flat and decentralized structures in research organizations correlate with higher productivity at the level of organizational units (Birnbaum, 1983). 2.3. Resources, recruitment and leadership Other important variables for productive research climate are human resources, instrumentation and funding. Pelz and Andrews (1966) report that actual resources are associated less highly with productivity than the resources researchers perceived they could access. Similarly, Andrews (1979) finds that the perceived accessibility of human resources but not the de facto level of human resources explains the largest amount of variance in a research unit’s performance. Recruiting outstanding scientists in a research team is another important variable. For example, Dill (1985) shows that highly productive research units can be distinguished by the significance they attach to hiring talent. The importance of recruitment also points to the influence of leadership. Hage (2006) argues that plural organizational leadership ensures diversity of research strategies and richness in ideas. The three directors of Institute Pasteur operated with diverse recruitment patterns but all three kept a look-out for creative people in their fields and then attempted to convince them to come to the Institute. However, leadership is crucial not only for recruitment, but also for directing research groups. Andrews (1979, p. 68, 102, 219) finds that effective leaders are involved in ongoing research. Active participation in the praxis of scientific work is important for leaders to understand the problems of the group, to motivate group members and to organize a coherent research program. This finding is also reflected in the literature review by Mumford et al. (2002) who suggest that leadership in creative environments requires predominantly technical and scientific expertise. In summary, the available literature provides several ideas which informed the development of our interview protocol (Section 4). However, opportunities abound for new work that probes. Much of the extant literature is based on historical analyses of science prior to recent developments in the structure and dynamics of scientific research in the advanced economies. Moreover, while there has been a significant focus on intra-organizational communication and the balancing of individual and institutional research goals, inter-organizational and institutional aspects have been somewhat less studied. For example, there has been little discussion of the role of research councils in sponsoring new research fields; there is little emphasis on how basic research activities are framed differently in public sector research organizations and private laboratories; and most studies are concerned with understanding productivity and recognition, but less with scientific creativity. In these respects, our study of creative scientific events and the research groups responsible for them seeks to generate new insights, particularly since we are embedding our cases in current organizational and institutional contexts. In the next section we introduce the methodology for identifying highly creative research accomplishments and the case study approach for examining work environments of highly creative scientists and their groups. 3. Methodology: identification of creative accomplishments and case study design The exploration of the features of the organizational and institutional context that have an effect on scientific creativity was carried out by means of a set of case studies anchored around selected individuals and their groups identified as highly creative in a new nomination method, previously reported in Heinze et al. (2007). Since the highly creative researchers were identified by independent experts and prize review panels, our work is based on ex post, external attributions of creativity by others rather than by modeling creative mechanisms at the individual level. 3.1. Identification of highly creative research accomplishments First of all, we conducted a survey that obtained more than 400 European and US nominations from 185 experts in two fields of research, human genetics and nanoscience/nanotechnology (referred to as nanotechnology), across five categories of prominence: highly cited researchers, active academic researchers, active industry researchers, journal editors, and research program managers.6 The two research fields were chosen to offer a comparison between a more established, disciplined-embedded field (human genetics) and an emerging interdisciplinary field (nanotechnology). Furthermore, both fields have undergone substantial growth in recent years (see Heinze, 2004 and You tie et al., 2008 for nanotechnology; Sulston and Ferry, 2002 for human genetics), while the science system as a whole is in a steadier state (Ziman, 1994). Field growth is an important variable for the development of creative ideas, because more novel ideas are produced, and the forces of sorting out original ideas are relatively weak (March, 2007, pp. 16–17). Therefore, it is especially fruitful to study the organizational context of creative research in growing research fields. In order to account for creative activity of several sorts, we introduced a typology of scientific creativity, with five stipulated categories and one open category for respondents to include types not included in the original list (Heinze et al., 2007; Table 1).7 In parallel with the nomination survey, we identified relevant prizes in the two fields, drawing on respondent nominations, other expert input, and our own knowledge. It should be noted that few prizes are specifically dedicated to nanotechnology, an exception being the Feynman Prize. Prizes for nanotechnology accomplishments are usually associated with a discipline (such as physics or materials research) or an organization (such as the Max- Planck-Society in Germany or the Centre National de la Recherche Scientifique in France). Our approach was therefore to identify relevant prizes broadly, then to carefully revie wall awards and laudations to explicitly identify relevant prize winners. We merged the nomination and prize winner data so as to offer a consolidated basis for studying creative research accomplishments (Table 2). Supplemented by additional web-based research, this provides us with a unique data source of information about creative research accomplishments in our two target science domains. We are particularly interested in scientists with multiple nominations, since recognition of their research is derived from more than one source. 3.2. Case study methodology Our database of highly creative scientists offers a foundation from which to develop case studies, since we can identify scientists by the number and type of creativity nominations, by field, and the character and timing of their creative research accomplishment.8 Hence, drawing on the subset of multiply-nominated creative scientists (i.e. those with most nominations from different sources), we undertook case studies of creative events of twenty research groups across Europe and the United States in the two fields of nanotechnology and human genetics. These cases explored the organizational and institutional dimensions of work environments in which creative research has been conducted. The theoretical framework for the case studies addresses the relationship between features of the context of research work and the occurrence of a creative event that has already happened and has been recognized by colleagues and other experts who participated in the nomination process. The theoretical question for the case study design is whether there is a predominant contextual pattern for creative events in scientific research. Since creativity is surely also a matter of individual talent, our approach cannot completely isolate the contextual factors from the individual abilities of the researchers as causes of the creative event. Rather, our exploration addresses a set of necessary conditions for a creative event to come to fruition given that the researchers in our cases are already regarded as highly talented. Moreover, the span of an individual’s research career is generally long enough for the same person to have worked in changing contextual conditions and it is therefore likely that the pursuit of research goals involves strategic calculations for the researcher in which assessment of contextual factors is inherent in choosing alternative paths of action. Therefore, it is possible to compare the effects of contextual factors at different times in the career of the same researcher. In order to obtain reliable data about the context from our informants, we used information from the nomination process, including publications, citations, prize citations, among other external indicators, to identify a specific creative event so that the features of the context could be related to actual conditions of work at a determined time and place. This significantly reduced the potential for recall problems, ambiguous statements and generic opinions in response to our interview questions because the creative events were prominent occurrences in the lives of our respondents, the circumstances of which they are likely to remember in detail. The cases were selected using parameters most relevant to features of the context, such as the research field (nanotechnology and human genetics), organizational affiliation (universities, medical facilities, industry R&D labs), geographical location which also provided diversity in various institutional and cultural dimensions (different regions of the United States and several European countries with their different funding mechanisms, academic styles, promotion rules, am ongother things), and different sorts of creative event as identified by nominators using our proposed typology. By conducting case studies under all these conditions we aimed at establishing whether there were emergent patterns that could confirm the presence of essentially the same sort of phenomenon across cases. In other words, and as a brief reminder, the logic of multiple case studies is not based on a representative sample for generalizing to a large population, as statistical inference does. Rather, cases are selected to elucidate the mechanism of a phenomenon for generalization to theory and concepts (George and Bennett, 2005; Eisenhardt, 1989). Each case included a fairly complete account, both historically and technically, of the creative event obtained from the key researcher (or researchers) associated with the accomplishment and validated by others familiar with the event (colleagues, collaborators, competitors or external observers). A comprehensive file with information on publications, co-authors, and citations, research themes, and the organizational context of the research group was also compiled. Then, in-depth interviews were conducted with the group leader, and follow-up interviews with colleagues and associates, group members, and other scientists who were knowledgeable about the circumstances under which the creative event materialized. In total, we conducted 44 interviews between November 2005 and February 2007 (see List of interviews). The interview protocol included questions related to the preparation phase prior to the creative accomplishment, the creative accomplishment phase, and factors related to research group, organizational and institutional levels. The main variables about which data is gathered are derived from the literature review above (Section 2). At the group level, variables include the size and composition of research team, communication patterns, quality of research leadership, and need and access to outside resources such as specialized equipment. At the organizational level, the variable set comprises (all variables at the time of the accomplishment): structure and size of the organization, centralization of decision-making, clarity of research goals, features of the funding arrangement, accountability burden, reputation and visibility of the organization. Finally, the institutional level variables include job mobility opportunities, competitiveness within the research field, and munificence of the funding environment. The case study method enables an in-depth analysis, capturing more of the texture and detail of behavior than is possible in conventional aggregate data-oriented methods. In addition, the case study method is “nonlinear” in the sense that the researcher learns from the case and, if appropriate, adjusts the focus of the research during the course of the project. Hence, there is no need to hold fast to hypotheses if they are clearly being discredited in favor of more accurate and valid explanations. Thus, the case study provides flexibility for the researcher to follow the most fruitful path. 3.3. Case description Of the twenty cases, ten were undertaken with researchers currently located in Europe, and ten with creative researchers currently located in the US. Fifteen cases were undertaken in nanotechnology, five in human genetics. There is some concentration in the creativity types 3 (new methodology) and 2 (empirical discoveries), followed by type 1 (theoretical concepts) and 5 (new synthesis). Thirteen out of twenty accomplishments can be characterized by more than one creativity type, and 1&3, 2&3 and 3&5 are the most frequent combinations in this regard. Five scientists/groups fall in the most select category: multiple nominations and multiple prizes. Seven cases are in the multiple nomination category, five in the prize and nomination category, and three cases in the multiple prizes category (Table 3). Fourteen cases appear as “serial” producers of creative accomplishments, indicating some institutionalization of effective practices for creative research. About half of the accomplishments have their roots in the mid-1980s, and three cases even in the late 1970s, indicating the substantial time necessary to move from idea generation to accomplishment. This time lag was due to resistance within the research community to accept these novel results and to incorporate them in the collective stock of knowledge. In one case, the group leader worked for about one decade solely on the problem until the research community accepted the novelty and usefulness of his work. In another case it took about nine years until an experimental result, that contradicted an established theory, could be matched with a new theoretical explanation provided by a collaboration partner. However, there are also several cases with rapid advancement from the preparation to the accomplishment phase, particularly when groups were involved in “priority races”. The creative events were accomplished across a wide range of institutional environments, predominantly in universities (N= 11), followed by basic research labs in industry (N= 4), settings in the public research sector spanning both a university and a government lab (N= 3), government labs (N= 1), and settings spanning government labs and hospitals (N= 1). The institutions in which our target scientists/groups work today are mostly universities (N= 15), but also mixed settings including hospitals and government labs (N= 4). Only one group with a creative accomplishment in an industrial basic research lab has remained in this institutional context. It should also be noted that US cases have some geographical concentration in areas which are already known to have a large share of R&D, such as the San Francisco Bay area and the greater Boston area. In contrast, we are not able to identify such concentration for Europe (Table 3). In the next section we report on the organizational and institutional factors that were obtained by systematic cross-case analysis. 4. Organizational and institutional factors influencing scientific creativity This section discusses key results from the twenty case studies with an emphasis on factors that support creative scientists and groups in their research, but also with findings regarding barriers to scientific creativity. The two levels: organizational and institutional are clearly intertwined, but we discuss them separately for analytical clarity. As noted above, our unit of analysis is the research group. The group is embedded in both an organizational and broader institutional environment which contributes to or constrains the capability of group leaders and group members to conduct research in a way that seems most fruitful to them. The comparison of the dimensions and variables from the case study protocol with the emerging dimensions demonstrates the learning process we underwent as we were exposed to the case study material. It shows how some of our initial expectations about the basic elements of the context of creativity were enriched and corrected by it. For example, spatial arrangements and infrastructure emerged as somewhat more important influences than anticipated. Furthermore, larger institutional developments, such as the severe budget cuts in (if not breakdown of) the former Soviet Union research system, and the increasing international mobility of scientists, figured more prominently in our data than initially expected. Finally, the comparatively strong representation of industrial R&D laboratories as organizational environments for scientific creativity, particularly until the early 1990s, corrected our initial expectation that the academic heartland is exclusively institutionalized in universities and government laboratories. In our policy conclusions, we will return to these and other issues. 4.1. Organizational level The cases indicate that creative researchers have a number of distinctive ways in which they manage their research groups. This includes highly effective supervisor–student relationships, the careful selection of new group members for complementary skills and attributes, and the flexibility to address new problems or ideas that arise. We also find that groups in our sample are relatively small at the time of the main creative event: typically around six to eight researchers, including juniors and students, and sometimes with only 2–3 researchers, but they benefit from organizational contexts that provide sufficient access to a relatively large variety of technical skills. A frequent factor associated with scientific accomplishments is stable research sponsorship, provided either through some kind of basic institutional funding or dedicated funding schemes for junior scientists. 4.1.1. Research autonomy First of all, among the prerequisites for a productive scientific atmosphere is a context in which there is a set of broadly defined problems or long term targets and carefully selected individuals are brought in as group members. They are then given freedom to pursue a more focused problem within the larger set of problems as a step toward the broad target. Freedom to define and pursue individual scientific interests within or beyond a broadly defined thematic area is central to understanding why scientists and their groups are highly creative. Individual scientific freedom is made even more productive when group members are conducting their work in units with many other bright and curious scientific minds who stimulate each other. Mutual curiosity and interest is a strong norm in all groups under consideration. For example, one group leader reported that one of his PhD students got upset because he expected the group leader to approach him about the ongoing experiment more frequently than just twice a week. Subsequently, the group leader communicated more frequently with his students. 4.1.2. Small group size Second, we identified small group size as important organizational dimension for the development of creative work. This confirms findings previously reported (Section 2). In fifteen of our twenty case studies, small group size was highly influential for the creative accomplishment; in another four case studies this variable had some influence. The analysis of the case studies indicates that research groups responsible for creative events often start with two people, the group leader and a PhD student or a post-doc. Later on, leaders deliberately limited their groups to no more than six to eight researchers (excluding technicians and other support staff). Small group size has a number of advantages. It allows the group leader to be actively involved in research and to stimulate effective scientific exchanges within the group. This is corroborated by the majority of our cases. In contrast, we have reason to believe that large research groups are less able than small groups to unleash the creative potential of their group members, because group leaders are forced to spend more time on administration than science which weakens the crucial link between group members and group leaders. In addition, small groups typically show a lack of hierarchical decision-making in relation to research activities. The flat structure of communication, with no difference in communication between formal hierarchical levels, fuelled the dynamics regarding creative research accomplishments. Furthermore, small group size fosters productive mentor–student relationships that larger groups have difficulty to establish and to maintain. However, several groups (although not all) did grow significantly in the period following the main creative event. This growth seems to be associated with following up and capitalizing on the opportunities that the creative event opened up. This raises an interesting question of the value of the research activities in each period, since the later ones would be deemed less creative but are critical for actually realizing the potential of the creative event itself. Also, it seems that group growth as a particular reward mechanism in science produces unintended negative effects, such as more hierarchical decision-making and reduced group leader involvement in research activities. 4.1.3. Complementary variety Third, the small groups were typically embedded in an organizational context that had a complementary variety of scientific skills and instrumentation. For example, one scientist reported his experience within an industry lab: “We were going to have lunch and of course if you come back from lunch with a thousand ideas because everybody is in a slightly different field but not so far from you so you can talk – it’s close enough. ”We also found collaboration within a university between theoreticians and experimentalists in physics that had enabled a highly recognized creative event in one of our cases. This type of environment provides numerous opportunities for stimulation, collaboration, the acquisition of new knowledge and research techniques or access to instrumentation. The combination of small work units in rich research contexts with requisite scientific variety allows for rapid elimination of dead ends when pursuing high-risk ideas. Researchers saw this as a critical factor since it allowed them to quickly test many of the possible paths to a solution for their problem and discard the ones that did not seem promising. Our case studies provided numerous examples of the importance to scientific creativity of a large, well-endowed organizational environment, with a good intellectual and technical infrastructure and access to a large diversity of skills and interdisciplinary knowledge across the organization. However, we found that the scientific diversity of an organizational environment alone may not foster creativity unless it is also linked to organizational arrangements that support multidisciplinary contact. These include spatial arrangements, such as the organization of laboratory facilities or office space, but also staircases or coffee bars designed to promote interaction. Social arrangements, such as lunchtime patterns, can also play an important role in fostering communication opportunities across departmental borders. In university contexts, for example, these interactions take place mostly across department boundaries. In our cases, some laboratories were more adept than others at facilitating these exchanges by cultivating a culture of shared resources and reduced bureaucratic requirements. For example, one group leader described the physical infrastructure of the university where the creative event materialized, as one in which all disciplines are united “under one roof”. Walking along the corridors initiated acquaintanceships and discussion between scientists from various disciplines. “Within three minutes I change from chemistry to physics to biology. When I walk to the electron microscope, I went through the faculties of physics and biology. These contact points are very important.” 4.1.4. Communication with groups in external organizations Fourth, effective communication with other groups that have complementary knowledge and expertise are an important factor for accomplishing creative events. For example, several theory groups depended on data from experimental groups in other research institutes, often abroad, or measurement groups needed access to sophisticated materials which they could neither produce in their own labs nor acquire from specialized companies. Interestingly, the emergent communication pattern showed that most of the in depth communication on matters close to the problem of interest to the group occurred with groups that were outside the organization; sometimes they were collaborators and on others competitors. On the other hand, the most important type of communication with groups within the organization was of a broader multidisciplinary nature and related to key skills the group itself did not possess. In other words, there is somewhat of an inverse relationship between cognitive distance and physical distance in the typical patterns of scientific communication. In addition to our case-study findings on communication across organizational boundaries, we examined the collaboration patterns of creative scientists in a quantitative way. Drawing on our database of multiply-nominated highly creative nanotechnology scientists, Heinze and Bauer (2007) find that these scientists collaborate much more frequently with other peers than scientists from a comparison group of similarly productive researchers; they have larger collaboration networks and more often link disconnected peers. Because of this particular communication pattern, creative scientists also publish in a wide range of academic journals, and thus they are capable of speaking to different audiences and specialties. Although, for technical reasons, Heinze and Bauer (2007) do not distinguish between intramural and extramural communication, the increasing gap in the volume of collaborations between the two types of scientists (i.e. creative researchers versus comparison group) demonstrates that getting timely access to complementary expertise, skills and instrumentation from other groups is important to achieving creative events. 4.1.5. Facilitating leadership Fifth, as we had anticipated from the literature review, both group and organizational leadership are important for the development of creative work. Effective group leaders perform many important roles. They bridge otherwise disconnected knowledge domains, carefully select new group members for their complementary skills and attributes, have the flexibility to address new problems or ideas that arise, help post-docs with a good idea both to attract funds so they can become self-sustaining and develop good intuition about the right measure of risk to take on with their new idea, and provide a protected area under which group members conduct their research. We also identified two types of mentor student relationships. In the first type, mentors provide a research avenue where their students can develop their particular research interests. For instance, mentors provide ideas and directions, but students arrange the experimental setting. In the second type, mentors recruit highly talented students in their group but do not set them on a particular research track. Their role is more responsive to the needs of the students. In addition to group leadership, we also witnessed the importance of directing research organizations through active leadership. In more than half of the cases, the director’s research vision at the R&D management level of the organization was influential in shaping the creative accomplishment. This vision is not so much about goal clarity as it is about goal-fruitfulness in generating more focused problems that are tractable and significant. Typical visions are “finding the highest possible storage density for computer memory” or “explain the Fragile-X syndrome ”which are more fruitful than clear, since the actual initial goal may undergo significant metamorphosis in the process of its pursuit. Perhaps most importantly, these leaders gave their junior staff freedom. Half of the creative events in our case studies are based on the research of junior scientists, highlighting the importance of providing independent research support to outstanding individuals at an early stage of their careers (see career stage in Table 3). As the focus of research missions is on the solution of a major problem and not on advancing disciplinary knowledge, this factor also has the attribute of encouraging interdisciplinary work because research teams were typically composed of people from a variety of disciplinary backgrounds that contributed to meeting the goals of the mission. 4.1.6. Flexible research funding Sixth, flexible research funds were found pivotal in several research breakthroughs in our set of cases. Flexibility means that funds are not earmarked for specific purposes; that group leaders have discretion about when and how to spend them; and that funds can be used for high risk/high potential investments. In this regard, flexibility allows scientists to shift research funds in the research direction that seems most fruitful to them. One group leader argued: “When someone has to invest or wants to invest unexpectedly much more money in a project, you need the flexibility. You need a good funding level in order to be able to afford flexibility .“In particular, core institutional funds, which are independent from success in attracting external grant money from research councils, have been found highly important to supporting scientific accomplishments in eight out of twenty cases, and in seven other cases these funds had some influence. Surprisingly, the four industrial research laboratory groups in our sample received high levels of such institutional core funding. For example, one group leader recalled that staff scientists in the industrial laboratory were not encouraged to write grant proposals, but to communicate directly with management about their demands for new resources. The interviewee argued that research managers who recruited him used to say: “I did not hire you to be a manager, I hired you because you are a scientist. I want you to do science. I want you to be in the lab.” In contrast, group leaders in universities were sometimes forced to raise funds for their research with small grants from many different agencies, and progress was achieved only because these group leaders used grants from research councils imaginatively. Another category of flexible funds are large, multi-year research awards provided to scientists in an ascending stage of their career. In total, six junior scientists were awarded prestigious and wellendowed individual awards. They were either supported by the Förderpreis of the Krupp von Bohlen und Halbach Foundation, the European Commission’s Young Investigator Award, the National Science Foundation’s Young Investigator Award, the James McDonnell Foundation’s Centennial Fellowship, or the Howard Hughes Institute’s Investigator Award. An in-depth study of these (and other) award schemes shows that they differ considerably with respect to target group and field, selection process and criteria, budget size, and funding duration; and that several of these schemes are powerful tools to support junior scientists (Heinze, 2008). In addition, we find unanimous consensus among our interviewees that too few such awards are currently available. 4.2. Institutional level So far, our findings suggest that there are several organizational factors that support the capabilities of research groups to accomplish creative scientific results. In addition, however, we have identified features in the institutional environment which facilitate or constrain the creative capabilities of research groups. In this section, we will report on such institutional factors. In brief, our cases indicate that job mobility is a necessary condition for creativity in science, because when scientists accept a new job, they tend to move to research units that offer an opportunity to change field or to address intrinsically risky research problems. Also, although competition is believed to be an important institutional mechanism in science, we observe several cases with little or no influence of competitive pressures in the institutional environment on the preparation or accomplishment phase of a creative event. Finally, we find that whilst the conservative procedures adopted by research funding agencies for allocating grants may be appropriate for “normal science” in established disciplines, they create many problems for scientists with original research ideas. 4.2.1. Job mobility We did not find that job mobility is as unidirectional a factor as one might assume. Several group leaders spent many years in the same place and either had been in one main institution their entire career or had made one major change in their career. The university setting has absorbed many researchers who moved away from industry labs when the attractive conditions in the latter deteriorated. But we did not find evidence of competitive recruiting mid-career as a mechanism to encourage creative research. Resources for hosting visitors or spending periods of time with other groups working on the same problem area had a greater effect on the success of the creative pursuits of our interviewees. We found that when they move, creative scientists tend to move to research units that offer an opportunity to change field or to address intrinsically risky research problems. Fundamental research labs of large, leading industrial companies were a magnet for such scientists, at least until the early 1990s. Other cases demonstrate that the United States has the most open academic job market in this regard and offers ample opportunities not only for scientists from Western European countries, but particularly for researchers from the former Soviet Union where the public research sector underwent severe budget cuts in the early 1990s. In eight of our cases “immigrant scientists” moved to different countries (including the United States, France, Japan, and Germany) permanently or for a long period in order to pursue their research. These researchers reported that they had to work much harder than native scientists in order to receive recognition, but all described their moves as pivotal for the development of their scientific skills, their success in accomplishing the creative event, and future career options. We conclude from these observations that job mobility is an important condition for an institutional environment that is conducive to creative research. 4.2.2. Reputational competition in the intellectual field Competition for reputation in the intellectual field seems to work in different ways depending on the phase of the creative work. In seven cases, competition was highly influential in either the creative event preparation or accomplishment phase. During the preparation phase, friendly competition within an organization is an important motivator. At the cusp, when an important result seems within grasp, chances are that groups in different organizations are also close to a significant achievement and the race to be first with important results will carry over even into the accomplishment phase, where important derivative results are pursued in a competitive fashion. For example, several groups were involved in “priority races” between competing scientists/groups. One group received materials for analysis and characterization from another group, but these materials were given, at the same time, to a rival group. In another case the priority race took place quite fiercely and without any mutual communication. In both cases, priority was given informally to the groups in our sample, but their publications appeared around the same time and even in the same journal as those of their rival peers. Yet, we also observed cases where competitive pressures in the institutional environment had little or no influence on the preparation or accomplishment phase. Most of the sewere found in the field of nanotechnology. One reason may be that the field of nanotechnology is still emerging: the new opportunities in nanotechnology have made possible many new research paths, such that several of our creative cases developed these with little initial competition, although still with substantial risk. Some of these paths have attracted considerable attention over time. One interviewee argued that in the preparation phase of the accomplishment, this new path was met with enthusiasm by the researchcommunity: “People liked it, right away. When engineers first sawit, itwas immediately a hit, there was ecstatic enthusiasm, it provided understanding, and we got inundated by request for talks and presentations.” Today, there are more than 50 groups worldwide working in this new area. 4.2.3. Funding agency behavior The manner in which responsibility for a certain field of research is allocated to a specific division of the funding agency and advised by experts in the area is a significant barrier for creative research because each division tends to award funds to scientists who have a record of publications in the area. Several group leaders had achieved their creative event on the basis of moving to a new field, or integrating new fields with their area of expertise. But one university group leader said her group “had no chance” to get money from a funding agency for their “wild ideas”. The group leader recalled that one always needs preliminary results in order to compete for external funds. Therefore, getting into a new field without having preliminary results is regarded as “almost impossible”. Another group leader argued that “field-hopping is bad for research grant income because it takes five years to build up credibility to get research funding”. The current research system does not appear flexible enough to accept that a scientist with an excellent track record in a given field can have the capability to investigate a phenomenon that involves moving into a new field and that there are synergies in funding such research. A second problem is connected to funding agencies jumping onto the bandwagon once the results of breakthroughs in research generate attention. Research councils and other funding agencies allocate funds for a program in the field and often set goals for work that is either already done, or unrealistic. Although programme officers in funding agencies may have a scientific background, they are perceived by several of our group leaders to have been out of research for too long to understand how research works. In consequence, the guidelines in calls for research proposals, according to one group leader, “tend to be wrong, and they do not present the actual priorities in the field”. Furthermore, many funding agencies require research proposals to set targets, or give exact details of the likely results, but this is often not possible with exploratory, open-ended research, characterized by one group leader as “a meandering path, you’re branching out, making new things all the time and closing up other things and so you’re moving through a difficult landscape to find your way to interesting things ”.We also found evidence that renewal funding is jeopardised if the expected results are not achieved. Several group leaders argued that a substantial portion of their research had not been described in any way in the research proposal. So, when it comes to grant renewals, funding agencies might argue that this research did not achieve its goals. Funding agencies also now require more accountability by scientists, and have increased the administrative burden on them. They require scientists to provide frequent progress reports, show they have worked the proposed hours or carried out the working steps according to the original proposal. Our case study results suggest that creativity would be promoted by having more flexibility in the use of grant income and less demands for constant progress reports. 4.3. Factor combinations The variables at the organizational and institutional level must be understood as interrelated contributing influences rather than as independent, cumulative factors. Set in its context, the creative research process that we were able to detect from interviews and supporting documents of the cases has mechanisms that combine many of the influences mentioned above in ways that are more than their simple aggregation. One of these mechanisms is found in large R&D laboratories in industry. Several of our highly creative researchers were recruited to these labs at an early stage in their careers, either as post-docs or junior staff researchers. They were then integrated into a mission-oriented research program while also allowed significant freedom to pursue an aspect of the overall program that most interested or excited them. This work environment was characterized by organizations that provided significant job stability for their staff scientists, a base level of funding, and access to a large diversity of skills and multidisciplinary knowledge across the organization. These research laboratories were well equipped with instruments and experimental capabilities that allowed the pursuit of empirical research in any direction the problem might suggest and had in-house, expert technicians to provide reliable experimental results in a relatively short period of time. The second mechanism is the university setting which is rather different from the industrial labs described above. We found that scientists that experienced their main creative events in a university context made efforts to create a setting as close as possible to the one described above while preserving the broader mission of academic work. The central difference between the main model described above and the academic setting, in the words of one of our interviewees, is that “industry labs are equipment rich and labour poor while universities are labour rich and equipment poor.” Therefore, university scientists must devote considerable efforts to gain access to the necessary equipment and compensate for the time demanded by graduate students, who are “in training mode” by selecting problems that are not too time sensitive. Three other important areas in which the academic setting departs from the fundamental research laboratory in industry mechanism are the conspicuous lack of core funding to protect against interruptions of the work, the burden of non-research academic obligations such as committee work and other service activities of the university, but also the strong individual freedom associated with a group leader’s position in the academic setting. In the next section, we review our findings in the light of the literature reviewed above, and we point to future directions for research. 5. Discussion Previous studies reported that intra-organizational communication across disciplinary or departmental boundaries is associated with a productive research climate (Andrews, 1979; Pelz and Andrews, 1966). Although in some cases this view is confirmed, we also found that extramural collaborations have a much greater benefit for scientific progress than was previously assumed. The changes in institutional and organizational conditions mentioned in Section 1, especially the encouragement of research collaboration, may explain the growing importance of extramural collaborations to scientific progress. However, timely access to skills and partners are not necessarily available within the boundaries of the research organization in which the focal group is embedded. As mentioned, there is an inverse relationship between cognitive and physical distance in the typical patterns of communication that facilitate the accomplishment of creative events. While accessibility to outside skills and resources tends to expand the capabilities of research teams to make rapid progress on matters close to the problem of interest to the group, other teams and resources within organizational boundaries provide a scientific reservoir for serendipitous observations generated through effective intramural communication. However, the opinion that deep knowledge and specialization at the individual level is integrated at the organizational level (Hollingsworth, 2002) is not fully supported. Creative scientists in our sample typically possess a rather broad scientific profile that distinguishes them frommore specialized normal scientists—a finding that is also corroborated by Heinze and Bauer (2007) with respect to the nanotechnology domain. In addition, several group leaders have accomplished creative events because they had changed their research field, for example from chemistry to optics, or from semiconductor physics to biophysics. In these cases, intellectual variety is integrated at the individual level rather than at the level of the entire research organization. However, our results do confirm previous findings regarding the strong correlation between small group size and per capita performance (von Tunzelmann et al., 2003). This is noteworthy because previous studies usually examined productivity and recognition, but not creative scientific accomplishments or research breakthroughs. Compared to the entire populations of researchers in the two fields, our twenty cases represent a very small fraction only, so the clear impact of small team size can be interpreted both as strong confirmation but also as an extension of previous evidence (because our dependent variable is creativity and not productivity). Interestingly, we obtained mixed findings with respect to departmental effects as identified by Long and McGinnis (1981) or Allison and Long (1990).Whereas the authors argue that once scientists have joined departments, their productivity levelmatches that of the department, in several cases of our immigrant group leaders the causality seemed to point in the opposite direction. Typically, these immigrant scientists had to prove their scientific capabilities by working much harder and by contributing substantially higher performance than native scientists before they were invited to join prestigious departments.Onthe other hand, group leaders in industry reported that the dynamics and pace within these fundamental laboratories was an important driver and an inspiration for their own creativity. Perhaps the two (strongly interconnected) variables where our findings diverge most sharply from previous evidence are funding and organizational leadership. Most importantly, in the public research sector the predominance of institutional block-grant funds (1st stream funding) has been replaced by a new regime based on competitive research council grants (2nd stream) and private foundation or industry sponsorship (3rd stream).Whereas previous studieswere concerned with the relationship between perceived or actual resource levels and performance (Andrews, 1979; Pelz and Andrews, 1966), our findings suggest that the continued expansion of peer-reviewed funding, in particular at early stages of the research process, may eliminate ideas that are judged by peers as speculative, unorthodox, or transdisciplinary. Peer-review criteria, such as plausibility and validity tend to encourage conformity, while originality draws upon and encourages dissent. For this reason, funding arrangements based on peer review tend to discriminate against the early stages of exploratory research, as they have an inherent tendency to support conventional mainstream research and scientific work that follows established research lines while ignoring visionary and high-risk approaches. Apart from a conservative bias, the double trend of decreasing institutional funding and increasing external sponsorship has at least two other consequences. First of all, winning funding competitions and reviewing increasing amounts of research proposals requires substantial time investment by scientists, time that they can neither spend on laboratory work and group interaction, nor for reading and contemplation. Since style and content of research proposals are different from presenting arguments and evidence in journal articles, these activities have reduced the precious time of the group leaders studied. Second, increasing extramural sponsorship requires a new type of organizational leadership. While research directors are expected to articulate a research vision, to recruit outstanding personnel, and tomotivate scientists (as argued in previous literature), a newtype of expectation has emerged: they need the capability to equip research organizations with appropriate funding from diverse sponsors and balance research budgets. Organizational leaders need to be successful in acquiring new grants and opening up additional funding channels. They must be competent in continuously monitoring the complex landscape of funding agencies and sponsorship programs. These newleadership role requirements are non-voluntary because organizations usually cannot afford to neglect their funding environment.9 Our case studies demonstrate, however, that not only university provosts and institute directors but increasingly group leaders are confronted with meeting these new roles. The consequences were described by one group leader, who had formerly worked in a fundamental industry laboratory, as follows: “When I came [to this university], I thought I would still do research. I haven’t done just one experiment in the seven years since I’ve been here, in the lab myself. Of course, I direct experiments but I don’t carry them outmyself. (. . .) People do most of my ideas but I’m a manager”. The discussion shows that while the institutional literature in science studies (mostly from the 1950s to the 1970s) offers useful starting point with respect to group and organizational variables, new themes have emerged that reflect the broader institutional changes in the research system since the 1970s. There are several examples where we believe our exploratory results could be fruitfully extended: (1) to understand the institutional forces that led to the marked decline of industrial companies funding exploratory and basic research (for an introduction see, for example, Chesbrough, 2003); (2) to explore the changed relationship between industry and public sector research in generating and diffusing knowledge (for an introduction see, for example, Evans, 2004); (3) to analyze quantitatively career trajectories of creative (entrepreneurial) group leaders (for an introduction see, for example, Levin and Stephan, 1991). More generally, it would be highly desirable to learn more about the differences between creativity in the natural and technical sciences on the one hand, and in the humanities and the social sciences on the other hand (for an introduction see, for example, Guetzkow et al., 2004). Also, we need more general theoretical propositions that serve as a framework for generalization and for stating cumulative hypotheses. Clearly, the renewed interest in organizational and institutional questions about the governance of research is an opportunity for more systematic theorizing. Several colleagues have started to investigate contingent variables at the group, organization and institutional level with organizational research outcomes (see, for example, Jordan, 2006; and the contributions in Jansen, 2007). Although a governance theory of research organizations does not yet exist, its development is underway. In the next and final section of this paper, we discuss conclusions for research management and research policy. 6. Policy conclusions This paper uses the results of a newnominationmethod to select a set of highly creative scientificaccomplishments inEurope and the United States in two scientific fields, nanotechnology and human genetics, in order to explore organizational and institutional factors associated with creative research. We employ a multiple-case study approach of twenty highly creative research accomplishments encompassing diverse types of scientific creativity. In these cases, we did not find remarkable differences by scientific field or creativity type; rather, the principal findings relate to mechanisms at the group, organizational, and institutional level. Several of these mechanisms are relevant for current policy making. In this section, we explore some of the lessons for research management and research policy. We collected evidence that a stimulating work context offers ample opportunities for fruitful scientific exchange, often across established cognitive domains. In contrast, the exploration mode is weakened when research groups are large and organized hierarchically. Despite this finding, large hierarchical groups can be found almost in every university and in every government laboratory. Large groups are either valued by research cultures (for example in Germany), or they emerge systematically because the predominant funding mechanism produces large structure after initial scientific successes (for example in the United Kingdom and theUnited States). Several of our case studies demonstrate that successful groups can grow substantially in a short time period (the Matthew effect, see Merton, 1957). Clearly, this mechanism is in conflict with the fact that breakthroughs are typically accomplished by small groups. Therefore, senior research management should be aware that highly creative research can be more difficult to undertake in large research groups, and that path-opening solutions in science seem to emerge more readily fromsmall research units. The size of research groups should be considered an important management objective for effective research governance, particularly in new and frontier research areas. Policy makers also need to think about newmechanisms that relieve successful scientists frommanaging toomany projects and too large groups, because large groups and hierarchical structures are barriers for creative research. Well-endowed research institutes with a good intellectual and material facilities infrastructure and access to a large diversity of skills and multidisciplinary knowledge provide numerous internal and external opportunities for stimulation, collaboration, the acquisition of new knowledge and research techniques, or access to instrumentation. Interactions within and across such organizational contexts are particularly fruitful when groups work in related and complementary fields of expertise and when research institutes have the requisite variety of skills and knowledge. Those involved in the planning of periodic reorganizations of research institutions should ensure that changes maintain and increase the breadth of disciplinary expertise available. Furthermore, the scientific diversity of research organizationsmay bemore likely to support creativity if linked to organizational, social and spatial arrangements that support planned and unplanned multidisciplinary contact. Organizational opportunities may include multi-disciplinary or cross-unit seed research awards, lab staff rotations, cross-training and inter-unit seminars and exchanges. Spatial arrangements, such as the allocation of offices, junior research space, hallways, coffee bars or laboratory facilities, and social arrangements, such as lunchtime patterns, may also be organized so as to encourage the opportunities for communication across departmental borders, between staff, regardless of their status, and between disciplines. Formore than three decades, the science system has been operating under “steady state” conditions (Ziman, 1994). Steady-state science has been accompanied by a decreasing lifespan for research projects, and this trend has given strength to the forces that eliminate unorthodox and original ideas. In several of our cases a certain kind of funding helped to reap the fruits of novel ideas: funding based on trust that scientists will do their work as well as they can. However, for many years, scientific research activity has been confronted with a high level of distrust, and this distrust is visible in the widespread use of performance indicators and by growth in measures such as evaluation, progress reports, management reports, audit certificates and the like. Permissive, trust-based funding does not appear to play the central role that it should play in the budgets of funding agencies. Therefore, policy makers should now be ready to accept that there is a clear need to provide appropriate funding for exploratory and high-risk research, even under the regime of steady-state science. When considering the institutional landscape that fosters creative research, it should be noted that a conspicuously favorable environmentwas found in a set of fundamentalR&Dlabs in industry that, due to changes in market conditions and industry strategies, has been drastically reduced in size over the last decade. Many of the industrial labs mentioned by our interviewees no longer exist and some that do no longer allow the sort of work that resulted in the creative events that earned them the recognition reflected in our nomination process. This raises another sort of policy question that is not limited to lessons for R&D management and grant mechanisms aimed at stimulating individual choices by researchers. Rather, it points to the overall direction of the innovation system and whether this change in research arrangements will have an effect on its capacity to be as creative as it has been to date. Acknowledgements This paper is based on research undertaken by the Project on Creativity Capabilities and the Promotion of Highly Innovative Research in Europe and the United States (CREA), sponsored by the European Union Newly Emerging Science and Technologies program (see: http://www.cherry.gatech.edu/crea/). Research assistance for the expert nomination survey and the development of data on scientific prizes was provided by Gerrit Bauer, Ajay Bhaskarabhatla, Taehyun Jung, Li Tang, Jue Wang, and Jingjing Zhang. We are grateful for survey responses supplied by numerous experts in Europe and the United States. An earlier version of this paper was presented at the 40th Anniversary Conference of Science and Technology Policy Research (SRPU), University of Sussex, Brighton, UK, 11–13 September, 2006. We appreciate the many helpful comments and suggestions provided by conference participants and by subsequent reviewers.

It is well attested that traditional academic labor usually involves teaching, research, and service (Barrow & Xu, 2022; Gunn et al., 2015). However, along with universities’ ever more intense research audits and their pursuit of research excellence and international rankings (Xu, 2023; Borg & Liu, 2013), research productivity has become one of the most influential elements in advancing academics’ professional development (Blackburn & Lawrence, 1995) and the international reputations of higher education institutions (Morze, Buinytska, & Smirnova, 2022). As the foundation and a critical predicting factor of research productivity, research engagement encompasses a range of activities, such as conducting research and publishing research findings, which empower teachers to become actively involved in research (Kyaw, 2021). Specifically, research engagement involves two distinct but interrelated practices—engagement in research, and engagement with research. The former refers to conducting research individually or collaboratively, while the latter refers to reading and using research (Borg, 2010; Kyaw, 2021). Rather than being a simple manifestation of publishing metrics, research engagement focuses on research innovation, knowledge contribution, the delivery of informed critique, and general community development (Parker, 2022), leading to teachers’ better understanding of their own professional practice and enabling them to gain professional knowledge about teaching and learning (Nasser-Abu Alhija & Majdob, 2017). Research engagement is also highly emphasized in Chinese university teachers’ professional development, since research outputs are usually weighted more heavily than teaching in staff evaluation and promotion decisions (Borg & Liu, 2013). In China, universities can be broadly divided into four types: “Project 985” universities, “Project 211” universities, regular universities, and colleges. The first two types are the most research-intensive and receive the most financial and other research support from the government. In contrast, regular universities often have an emerging research culture, and colleges are almost always wholly teaching-focused. Nevertheless, various national policies have been issued in China to encourage university teachers to be research-engaged and to publish their work, such as Suggestions on Strengthening the New Era of Education Scientific Research Work (Ministry of Education of the People’s Republic of China, 2019). This stresses the responsibility of higher education teachers (or academics, framed in other geographic contexts) to teach and research, and demands adequate support from institutions to guarantee their research. However, during the process of encouraging teachers’ research engagement, a number of problems have arisen. For instance, publishing in local Chinese journals is highly competitive. During the period 2021–2022 there were only 16 CSSCI (Chinese Social Sciences Citation Index) journals available for around 700,000 language teachers in China to publish their research. Hence, increasing numbers of language teachers are trying to publish in international journals. However, many lack confidence in their academic English proficiency, especially when they are competing with scholars who are native English speakers (Zheng & Gao, 2016). In general, then, Chinese teachers’ research engagement and research outputs are unsatisfactory, hindering their personal and professional development (Bai, 2018). Nevertheless, despite the significance of research engagement, we know little about how to promote teachers’ research engagement, and especially about how, and the extent to which, different factors influence it (Heng, 2020). To address these issues, we investigate the influence of individual and institutional factors (e.g., motivation, self-efficacy, institutional support) on Chinese university EFL teachers’ research engagement through the theoretical constructs drawn from Tien et al.’s (2019) Research Capacity Model (RCM) and relevant literature. 2. Theoretical framework We first adapted and adopted Tien et al.’s (2019) Research Capacity Model (RCM) (see Fig. 1) to explore the relationship between teachers’ research engagement and a variety of individual and institutional factors. 2.1. Research capacity model Tien et al.’s (2019) RCM emphasizes the importance of considering an academic’s research capacity in the context of their professional practice. The model offers an explanatory framework for the relationships between teachers’ research capacity and a range of influential factors, particularly unobservable internal influences and visible external influences on their research capacity. Each level of the model represents a perspective from which to consider and understand an academic’s research capacity. As shown in Fig. 1, the model comprises five levels, with research capacity at the core and a range of individual, institutional, communal, and societal factors in the outer layers. These layers are understood to be interdependent, and to interact with one another (Tien, Hai, & Viet, 2019, pp. 459–469). In other words, any change at the core of the model—i.e., academics’ research capacity—depends largely on the impacts of the factors in the outer spheres. 2.2. Utilization of the model This model is a suitable theoretical framework for the current study because it may offer unique insights into how factors at multiple levels interact and influence teachers’ research engagement, which is seen as the key to research capacity (Tien et al., 2019). In the present study we further divide research engagement into engagement in and with research. Moreover, we mainly focus on the first two levels of the model, containing individual and institutional factors, because they have been shown to exert more influential and direct effects on teachers’ research engagement and capacity (Ajjawi, Crampton, & Rees, 2018; Tien et al., 2019; Uwizeye et al., 2021). According to Tien et al. (2019), the individual level mainly involves the factors that motivate teachers to be research-engaged and how they view their current capabilities to engage with research, thus pointing to factors related to research motivation and self-efficacy. Alongside these individual-level factors, Tien et al.’s (2019) model also depicts how teachers’ research engagement (capacity) may respond to factors from the external environment, especially the most immediate one, namely their institutions. To a large extent, teachers’ research engagement—their behavior of conducting and/or using research—depends simultaneously on individual factors (e.g., intrinsic and extrinsic motivation, self-efficacy) and institutional factors (e.g., institutional support, networks, national policies). Fig. 2 shows the adapted Research Engagement Model we further developed by integrating relevant literature as illustrated in the following section, and applied to the current study as our theoretical framework. 3. Literature review 3.1. Research engagement Although a large number of studies have been undertaken to discover the factors that stimulate research activities and promote teachers’ research engagement and productivity, there are still some gaps (e.g., Blackburn & Lawrence, 1995; Bland, Center, Finstad, Risbey, & Staples, 2005; Heng, Hamid, & Khan, 2020). First, there is growing interest in how often teachers conduct research, as evidenced by the broad literature coverage (e.g., Borg, 2009; Borg & Liu, 2013; Sadeghi & Abutorabi, 2017). However, it is important to note that the findings regarding research engagement among teachers vary across different contexts. Therefore, further investigation is necessary to thoroughly understand the specific circumstances that contextualize teacher research engagement, particularly in jurisdictions like China, where relevant research is limited (Wang, Zhang, & Zhang, 2020). Second, although frequencies of and barriers to doing research have been investigated widely (e.g., Borg, 2009; Sadeghi & Abutorabi, 2017; Sakarkaya, 2022), teachers’ research engagement and factors facilitating it have rarely been studied, suggesting a need for further research. Third, while most of the existing literature investigates teachers’ needs, benefits, problems, and possible solutions at the theoretical level, there is a lack of empirical studies exploring their actual research engagement, or the extent to which they engage with/in research (Mehranirad & Behzadpoor, 2022). Finally, some scholars have identified academic qualifications, motivations, gender, and research self-efficacy as the most important individual factors related to research (Mantikayan & Abdulgani, 2018; Uwizeye et al., 2021), but there is a paucity of systematic studies of how individual and institutional factors influence English language teachers’ research engagement (Farsani & Babii, 2019). Therefore, informed by Tien et al.’s (2019) RCM and in order to develop a deeper understanding of teachers’ research engagement, we focus on exploring the influences of individual aspects, specifically research motivation and self-efficacy, as well as the impact of institutional support on Chinese university EFL teachers’ engagement in and with research. 3.2. Motivation Motivation can be defined as the dynamic and evolving accumulation of arousal within an individual. It serves as the driving force that “initiates, directs, coordinates, enhances, concludes, and evaluates the cognitive and motor processes” involved in selecting, prioritizing, operationalizing, and ultimately acting upon one’s initial wishes and desires, either with success or failure (D¨ornyei & Ott´o, 1998, p. 65). Motivation can be categorized into two types: intrinsic motivation and extrinsic motivation. Intrinsic motivation is driven by the inherent interest and satisfaction derived from engaging in an activity, acting as the primary reasons that attract individuals to participate. In contrast, extrinsic motivation involves external incentives or pressures that entice individuals to pursue a particular activity (Reeve, 1995). Previous investigations of what motivates teachers to do research have not reached a consensus (e.g., Li & Zhang, 2022; Bai & Millwater, 2011; Borg, 2009; Borg & Liu, 2013; Yuan, Sun, & Teng, 2016). For example, some scholars have found that English language teachers tend to highlight internal factors when explaining their research motivation (Borg, 2009), but Spanish language teachers rely more on extrinsic motivation to become research-engaged (Gironzetti & Mu˜noz-Basols, 2022, pp. 1–28). In Chinese higher education, Chen, Gupta, and Hoshower (2006) identified that both intrinsic motivation (peer recognition, respect from students, and satisfaction) and extrinsic motivation (tenure, pay raises, professorship, administrative assignment, reduced teaching load, and promotion) influence teachers’ research. However, rather than a combination of both, some scholars found that intrinsic motivation is more influential in encouraging Chinese language teachers to do research (Bai & Millwater, 2011). However, others found contrasting results, suggesting that extrinsic motivation (e.g., improving teaching and fulfilling employer’s requirements) plays a more significant role in driving language teachers to conduct research (Borg & Liu, 2013). Even though the findings of these studies are inconsistent, likely due to their varying research designs and contexts, they all reveal the potential effects of intrinsic and extrinsic motivation on teachers’ research engagement, and therefore are worthy of attention in order to develop more contextualized and systematic research. Furthermore, this research imperative is reinforced by the establishment of the Academic Committee at the tertiary level to evaluate teachers’ academic achievement, based on the Higher Education Law of the People’s Republic of China (1999). Since the Academic Committee was formed Chinese universities have adopted different means to promote teachers’ research motivation, such as bonuses, monetary or non-monetary rewards, and teachers have been either pressured or motivated to undertake research (Dai, 2008). 3.3. Research self-efficacy Research self-efficacy can be defined as an individual’s belief in their own capability to conduct research in a competent and effective manner (Forester, Kahn, & Hesson-McInnis, 2004). Research self-efficacy plays a vital role in faculty members’ engagement in research activities in the university context (Hemmings, Kay, Sharp, & Taylor, 2012); scholars have found that teachers with high research self-efficacy are more productive than those with low research self-efficacy (Pasupathy & Siwatu, 2011). It has also been argued that high self-efficacy teachers can keep calm when facing difficulties, whereas low self-efficacy teachers often feel out of control, and their negative emotions can further hinder their ability to solve problems in research (Harun, Putrawan, & Miarsyah, 2019). Research has also shown that Chinese language teachers have weak research capacity (Wen, 2020) and generally lack confidence in conducting research (Kumaravadivelu, 2012). Wang (2017) proposed two possible reasons for this. Firstly, as foreign languages fall within the academic field of foreign studies, they are neither urgently needed for scientific innovation and industrial upgrading, nor essential for the spread of traditional culture. Consequently, Chinese EFL teachers often feel undervalued and have little confidence in undertaking research. Secondly, because they are affiliated with an imported discipline, EFL teachers need to consider both international standards and Chinese cultural constructions, which makes it hard for them to be identified by both sides and weakens their self-efficacy in terms of undertaking research (Liu, Kong, & Chen, 2019; Wang, 2017). Given the significance of and problems related to teachers’ research self-efficacy, aspects such as how teachers gain practical knowledge and how self-efficacy belief positively affects planning, conducting, analyzing, presenting, and writing up research are certainly worthy of further exploration (Jang & Shin, 2011; Wyatt & Dikilitas¸, 2016). 3.4. Institutional support Research support may be defined as the provision of resources that can boost a faculty member’s ability to engage in scholarship (McGill & Settle, 2012). This involves a proper organizational work environment with effective policies, a reasonable structure, and supportive resources for the job (including incentives, practical goals, skills, and staffing), all of which are necessary for any significant research (Bland et al., 2005). A large body of literature identifies how different types of institutional support affect teachers’ research engagement. For example, Borg (2009) postulated that organizational and institutional factors, such as time support, research training, emotional support, etc., significantly promote teachers’ research engagement. Likewise, some scholars have reported a positive relationship between teachers’ research productivity and infrastructure support, such as funding, technical assistance, and staff support (Gruppen, Frohna, Anderson, & Lowe, 2003; McGill & Settle, 2012). In the United States, Randazzo, Priefer, and Pasupathy (2021) identified institutional rewards and opportunities, mentoring or sponsorship, organizational leadership, and generative behavior (independent sponsorship and socialization of newcomers) as positively related to research and publication activity; however, they also reported that a disconnect between an institution’s rhetoric and its actual actions affected teachers’ engagement and motivation for research. It appears that the more institutional support the institution provides, the more teachers can become involved in research (Shamai & Kfir, 2002). However, as well as defining the different types of institutional support, previous research also suggests that teachers’ perceptions of research engagement are highly context-specific (Sippel & Sato, 2022) and there is a lack of institutional support in various contexts. Sato and Loewen (2019) found that limited institutional support hindered teachers’ research engagement in Chile. This result is supported by evidence retrieved from Spanish language teachers (Gironzetti & Mu˜noz-Basols, 2022) and Myanmar teacher educators (Kyaw, 2022). In China, Luo and Hyland (2016) found that insufficient institutional support was one of the main reasons underlying university teachers’ difficulties in publishing manuscripts and gaining citation counts. Although the relationship between research engagement and institutional support has been widely studied in many other countries, such as the United Kingdom, the USA, Canada, Australia, New Zealand, Israel, and Bangladesh (e.g., Anwaruddin & Nasrin, 2015; Borg, 2010; Hemsley-Brown & Sharp, 2003), it is still an under-researched topic, especially in the Chinese context (Dundar & Lewis, 1998; Sato, Loewen, & Pastushenkov, 2022). After reviewing the literature, we further developed our theoretical framework—the Research Engagement Model (REM)—by integrating sub-factors drawn from the literature (see Fig. 3). Moreover, the above studies suggest an imperative need to investigate how different factors, especially teachers’ extrinsic and intrinsic research motivations, research self-efficacy, and institutional support, can influence their research engagement in a particular context, making the investigation situated and contextualized. Therefore, drawing from previous research, we adopt the perspective of the RCM (Tien et al., 2019) and propose the following research questions. 1. How do Chinese EFL teachers’ research motivation, self-efficacy, and institutional support influence their research engagement? 2. Which factors are the most influential in affecting Chinese EFL teachers’ research engagement? 4. Methodology 4.1. Instrument Based on the REM developed from the RCM (Tien et al., 2019) and other relevant literature, we designed a self-reported questionnaire to investigate how motivation, self-efficacy, and institutional support, with their sub-factors, influenced Chinese EFL teachers’ research engagement. A six-point Likert scale was adopted because an uneven number of options may encourage respondents to choose the middle category (e.g., not sure) and avoid making a definitive choice (D¨ornyei & Taguchi, 2010). The questionnaire consists of four sub-questionnaires: 1) Questionnaire on Teacher Research Motivation (QTRM), examining intrinsic and extrinsic motivation. Specifically, intrinsic motivation was measured by teachers’ interest in research, responsibility for conducting research, mastery of research skills, sense of achievement, and flexibility of research. Extrinsic motivation included respect from others (e.g., colleagues, students) and research compensation (e.g., promotion, salary raise); 2) Questionnaire on Teacher Research Self-efficacy (QTRSE), examining teachers’ self-efficacy, or their confidence, in using research design skills, practical research skills (skills related to actually conducting research, such as data analysis) and writing skills; 3) Questionnaire on Institutional Support for Teacher Engagement (QTRE), collecting teachers’ demographic information and examining teachers’ engagement in and with research. The scales and items used in the first three questionnaires were selected from previously established instruments (Amabile, Hill, Hennessey, & Tighe, 1994; Angaiz, 2015; Phillips & Russell, 1994), whereas the fourth questionnaire was developed by the authors of this study. The validation of the above questionnaires has been published in previous work (Li & Zhang, 2022). It was noted that the four sub-questionnaires were initially designed in English but translated into Chinese in the end to help participants understand the content better. 4.2. Participants and data collection This study focused on full-time EFL teachers who teach English in Chinese universities. As full-time staff, these teachers must undertake both teaching and research responsibilities as stipulated in Chinese Higher Educational Law. The participants were invited to complete the questionnaire within four weeks, starting from the beginning of the first semester of the 2020–2021 academic year. The data collection instruments were anonymous user-friendly online questionnaires for EFL teachers, administered via the internet including popular Chinese social media platforms such as WeChat and QQ. During the data collection stage snowball sampling (D¨ornyei, 2010) was employed to recruit participants, because the membership of the groups being studied was “not readily identifiable” (p. 61), and it was difficult to access suitable group members (D¨ornyei, 2007). Using our own social networks we emailed five Deans of Foreign Language Schools from one 985 university, two 211 universities, two regular universities, and one college, inviting them to take part in this research and to recommend teachers in their schools who met the research criteria and might be willing to join in this research. In turn, these teachers then recommended other potential teachers. In this way, we collected 536 questionnaires in total. 4.3. Data analysis Before data analysis, missing values in the questionnaires were identified. To clean the database, frequency, maximum/minimum scores, means, and standard deviations were calculated to screen the data for missing or incorrect values. Mardia’s Mahalanobis d2 values were utilized to check the normality of the data, as well as to check for outliers. As a result, 13 questionnaires were excluded for having straight-line answers, and 15 questionnaires were removed because of missing values. The total proportion of questionnaires with missing values was about 5%, which was less than the cutoff value of 10% at this stage (Field, 2018). We were left with a final total of 508 questionnaires suitable for data analysis. Statistical models are efficient and convenient ways to describe the latent structure underlying a series of observed variables (Byrne, 2016). Therefore, the factorial structure of each construct within the hypothesized model was tested respectively, using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to specify the relationships of the variables within the measurement model. The purpose of the measurement test was to evaluate whether the observed variables accurately reflect the desired factors, and to determine the extent to which the measurement model fits the empirical data. Each measurement construct was examined separately with the sample of 508 Chinese EFL teachers who participated in the online survey. The model fit evaluation was conducted in line with multiple model fit indices. The final choice of well-fitting model for the measurement constructs was specified by the model fit (for details please refer to Li & Zhang, 2022). Structural Equation Modeling (SEM), a multivariate statistical framework to model complex relationships between directly and indirectly observed variables (Stein, Morris, & Nock, 2012), can be used to describe the relationships among various measurement model components. It can also address research questions associated with complex causal relationships among latent constructs (Nusair & Hua, 2010). In the present study we conducted SEM analysis to investigate the predictive effects of teachers’ research motivation, self-efficacy, and institutional support on their research engagement. 5. Results This study recruited a total of 536 Chinese university EFL teachers on a voluntary basis. Table 1 displays the participant information. Roughly half of the teachers (49.2%) in this study were recruited from regular universities. The number of teachers from universities designated as “985” and “211” was nearly equal, each contributing close to one fifth of the total participants. In contrast, college teachers comprised the smallest portion, accounting for only 8.9% of the teachers. Notably, there were significantly more female teachers (380) than male teachers (128), almost triple the number of male participants. The largest age group among the teachers was those aged between 41 and 50 years, constituting almost 40% (199 teachers) of the total sample. Additionally, a similar percentage of teachers fell within the age range of 31–40 years. Those under 30 years represented 13.2% (67 teachers) of the participants, while 8.9% (45 teachers) were over 51 years old. When considering academic ranks, lecturers and associate professors dominated the study, accounting for over four fifths (82.9%) of the total teachers. Assistant lecturers and full professors were less well represented, making up 11.6% (59 teachers) and 5.5% (28 teachers) of the participants, respectively. 5.1. Structural Equation Modeling (SEM) with full sample The conjoint SEM structural model was tested after establishing the four CFA measurement models. The results are as follows. Based on the theoretical framework, the Research Engagement Model (REM) was tested. The three measurement models (research motivation, self-efficacy, and institutional support) were used to predict their relationship with teachers’ research engagement. Table 2 shows the model fit indices of SEM for research engagement, which were all in the good range (χ2 = 1789.461; χ2/df = 2.206; CFI = 0.95; RMSEA = 0.049; SRMR = 0.0443; gamma hat = 0.92, TLI = 0.95). Fig. 4 shows a simplified graphic representation of the SEM model only with significant paths. Over the total of 12 paths shown in Fig. 4, three factors (mentorship, working environment, and compensation) significantly impacted engagement with research. However, no factor directly affected engagement in research. Two factors under institutional support had a positive impact on engagement with research, namely mentorship (β = 0.145, p˂ 0.05) and working environment (β = 0.168, p˂ 0.05). This finding suggests that teachers’ engagement with research was raised when mentorship was provided and the working environment was improved. In terms of extrinsic motivation, compensation had a negative effect on engagement with research (β = 􀀀 0.238, p˂ 0.05), suggesting that compensation settings posed a hindrance to the level of research engagement among teachers. A total of 40.8% of their engagement in research was explained by this model (R2 = 0.408, f2 = 0.65, p˂ 0.001), while 39.1% of their engagement with research was explained (R2 = 0.391, f2 = 0.65, p˂ 0.001). The regression weights are shown in Table 3. 6. Discussion The Research Engagement Model (REM) that we developed from the Research Capacity Model (RCM) (Tien et al., 2019) and relevant literature has been evidenced to be appropriate to explore the influence of individual and institutional factors on EFL teachers’ research engagement—engagement in research, characterized by their conducting research individually or collaboratively; and engagement with research, demonstrated by their reading and using research (Borg, 2010; Kyaw, 2021). The results suggest that none of our factors significantly influenced Chinese university EFL teachers’ engagement in research. This finding may be attributed to the fact that the majority of our participants are female teachers (n = 380, 74.8%), reflecting the prevalent gender composition of EFL teachers in China. However, compared to male teachers, female teachers often exhibit lower academic aspirations (Heng et al., 2020) and tend to engage less in research as the principal or co-investigator of research projects (Papanastasiou & Karagiorgi, 2019). These disparities may derive from the traditional social and institutional discourses that often assign females more secondary and supporting roles (Heng et al., 2020). The results also reveal that intrinsic motivation did not significantly influence teachers’ engagement with research. This suggests that Chinese English teachers have little interest in and internal drive to do research. One possible explanation for this is that the competition for publishing articles in recognized journals, i.e., CSSCI- and SSCI-indexed journals, is intense, making it almost a mission Fig. 4. Research Engagement Model for the Full Sample Note. Mentorship = Mentorship Support, Working Environment = Working Environment of Teachers; Design = Research Design Skills, Practical = Practical Research Skills; Writing = Writing Skills; Interest = Interest in Research, Responsibility = Responsibility for Conducting Research, Skill = Mastery of Research Skills, Achievement = Sense of Achievement, Flexibility = Flexibility of Research, Respect = Respect from Others, Compensation = Compensation for Research (e.g., promotion, salary raise). impossible for most EFL teachers in China. For one thing, there are only 16 CSSCI journals in China publishing articles in the broad discipline of linguistics. For another, many teachers may feel incapable of competing with native English writers to publish in international journals (Zheng & Gao, 2016), although whether non-native English writers are disadvantaged in international publishing is a controversial issue (Hyland, 2016; Politzer-Ahles, Holliday, Girolamo, Spychalska, & Berkson, 2016). This bumpy and difficult publishing experience may erode teachers’ interest and drive to be research-engaged. In the end, research without any research outputs (e.g., publications, grants) is often not recognized by universities in staff evaluation and promotion (Barrow & Xu, 2022), thus bringing the teachers no sense of achievement. In addition, most of our study participants come from regular rather than research-intensive universities. In such institutions, with a heavy teaching workload and only an emerging research culture, many may lack research time and skills. Therefore, taken together, the challenging nature of research and publishing, the lack of time and relevant skills have contributed to their low intrinsic motivation for research engagement. In contrast to the English language teachers’ lack of intrinsic motivation to engage with research, our analysis shows a negative relationship between compensation (extrinsic motivation) and engagement with research. In Chinese universities, annual performance reviews are widely used for staff evaluation and promotion, and these reviews usually reward those who over-perform in research through financial compensation. However, similar to our results, Borg and Liu (2013) also found that this kind of evaluation and compensation appeared to function negatively in promoting staff research engagement. This negative correlation may be explained by the fact that the compensation is related to research outputs rather than research engagement, and such external policies to a large extent encourage competition rather than professional growth and thus are destructive to teachers’ research engagement (Kyaw, 2022). Moreover, research engagement does not necessarily lead to research outputs, which may involve many other factors and procedures (Xu, 2022). In other words, research engagement demands effort and time, which could all be in vain and receive no recognition from the institution (e.g., in the form of promotion, salary raise, or bonus) if no paper is published or no grant funding is obtained as a result of the engagement. After a cost-benefit evaluation, teachers may opt to divert their efforts toward other activities which offer more secure financial and social gains (Horodnic & Zait¸, 2015). The quantitative results reveal no relationship between teachers’ research self-efficacy and their research engagement. In other words, Chinese EFL teachers’ self-efficacy regarding their confidence in research design skills, practical research skills (e.g., data collection, data analysis), and writing skills were non-significant in terms of predicting their research engagement. This finding supports Pasupathy and Oginga’s (2014) study showing a weak relationship between self-efficacy and research engagement. A possible explanation for this is that Chinese EFL teachers have inadequate knowledge about research and research writing, since many hold only a Master’s degree and have limited research training and little research and publishing experience. Nevertheless, previous training and research experience are reported to positively predict language teachers’ engagement with research (Sippel & Sato, 2022). In this sense, the participants’ research design skills, practical research skills, and writing skills may be insufficient for them to be research-engaged. However, it is also possible that a more sensitive instrument of research self-efficacy, for example, with more options, may generate more insights regarding Chinese EFL teachers’ research engagement. Analysis of the QISTR indicates that institutional support in relation to mentorship and the teachers’ working environment is the most influential factor that positively correlates to teachers’ engagement with research. Previous studies have already revealed the importance of mentorship, showing that adequate mentor support from institutions can increase faculty members’ research endeavors (Borg, 2006; Eby & Robertson, 2020). Experienced teachers may mentor novice teachers and those with weak research foundation by way of offering suggestions and sharing research experience, including stories of successes and of dealing with challenges and emotional turbulence, and by inviting them to be part of the research projects to gain first-hand research experience and skills (Randazzo et al., 2021), or via collaboration between experienced and inexperienced teachers in co-authorship (Nguyen, Bui, Nguyen, Tran, & Tran, 2021; Owan, Ameh, & Anam, 2023). With the help of suggestions, shared experience, and collaborations, teachers who are weak in research could build their research skills and become more confident and motivated in engaging with research. Therefore, establishing and maintaining a research mentorship community with institutional support may help to create a learning and modeling space where EFL teachers can gain access to knowledge, experience, and resources for conducting research, engaging with research activities, and handling negative emotions and other problems and issues while researching. The results also suggest that a sound working environment (e.g., reasonable workload, funding, training, learning resources, good office conditions) could significantly promote teachers’ research engagement by creating a relatively relaxing and supportive space to concentrate on research. According to Way, Morgan, Larremore, and Clauset (2019), the drive for early-career faculty to be research-engaged involves the environment where they work rather than the environment where they were trained. An appropriate working environment guarantees auxiliary support for research, which would make teachers more willing to be research-engaged. 7. Conclusions and limitations By offering a contextualized understanding of how individual and institutional factors influence Chinese EFL teachers’ engagement in and with research, this study contributes to the field in several ways. Theoretically, by drawing on Tien et al.’s (2019) Research Capacity Model (RCM) and the previous literature, we develop the Research Engagement Model (REM) that zooms in the concept of research engagement in the individual and institutional spheres, distinguishing between engagement in research and engagement with research. This distinction deconstructs the notion of research engagement, providing researchers and teachers with a means to understand and analyze teachers’ research engagement at more granular levels. Methodologically, we employ Structural Equation Modeling (SEM) to examine the relationship between Chinese university EFL teachers’ research motivation, self-efficacy, institutional support, and research engagement. This method offers a new perspective revealing the extent to which individual and institutional factors intertwine to affect teachers’ research engagement, and thereby sheds light on EFL teachers’ personal and professional development related to research. Practically, by revealing such relationships, our study unveils a low or unsatisfactory level of EFL teachers’ research engagement in that many lack internal and external motivations to read or use research (engagement with research), letting alone conducting research solely or collaboratively with complex research skills (engagement in research). Therefore, on the one hand, it is vital for institutions to stimulate teachers’ research interest and confidence by fostering a supportive and collaborative research culture, thereby providing teachers with time and space to access various research resources (e.g., research mentorship, periodic training sessions, drop-in research consultations), and opportunities to attend academic conferences and be involved in different levels of research collaboration—the breeding place of new knowledge (Fauzi, 2023). Such institutional support can be particularly critical for female EFL teachers, who often juggle multiple family roles alongside their academic responsibilities and are sometimes perceived as secondary academic citizens (Heng et al., 2020; Papanastasiou & Karagiorgi, 2019). On the other hand, EFL teachers should also be encouraged to use diverse free online resources (e.g., learning platforms, research workshops) to improve their research skills. We also call for emotional support from institutions to help teachers handle negative emotions (e.g., pressures, frustrations, self-doubts) resulting from the challenging research and publication process. We also acknowledge several limitations of the study. By using snowball sampling, this large-scale study recruited sufficient participants. However, this method does not lend itself to generality. For future studies, we suggest using a random sampling strategy to explore EFL teachers’ research engagement with a more representative sample of participants. Qualitative methods may also be useful to offer valuable insights into the profound reasons behind the relationships revealed by this study. Furthermore, factors at the community and societal levels may also be considered in order to develop a more holistic understanding of teachers’ research engagement.

Research collaboration is one of the key social features of modern science (Milojevic´ 2010; Gla¨nzel and Schubert 2005). It can be observed and studied at many levels: individual, institutional, national, disciplinary, etc. At all of these levels, major research questions are how research collaboration is structured, how it evolves, and how it is related to research productivity and the impact of multi-authored publications. As emphasized by Gla¨nzel and Schubert (2005), research collaboration can be reliably captured by so called co-authorship networks since co-authorship is one of the most concrete and well documented manifestations of collaboration in science. The nodes in a co-authorship network represent researchers. Two researchers are connected by an undirected link if they co-authored at least one publication together. Additionally, weights can be assigned to links in order to express the strength of research collaboration. Empirical studies investigating scientific collaboration date back to the 1960s. A more recent resurgence of interest in the field was sparked by the observation of the small-world and scale-free phenomena in various types of complex, real-world networks (Watts and Strogatz 1998; Barabasi and Albert 1999). Mark Newman in his seminal papers (Newman 2001a, b) proposed a general methodological framework based on metrics and methods of complex network theory to analyze scientific collaboration networks. Following his framework researchers investigated the structure and evolution of a variety of co-authorship networks observing properties such as: • heavy-tailed distributions of node centrality metrics, • network evolution governed by the preferential attachment principle, • the small-world property additionally emphasized by shrinking diameters, • an evolutionary increase of average node degree and densification laws, • the funneling effect, • the existence of a giant connected component, • assortative mixing patterns, and • a high degree of local clustering and community structures at the mesoscopic level (Savic´ 2015). A vast majority of existing studies on co-authorship networks are mainly oriented towards inter-institutional, international, national and collaboration within scientific disciplines (Kumar 2015; Savic´ 2015). To the contrary, empirical studies of intra-institutional research collaboration networks are exceptionally rare. In a recent comprehensive literature review made by Kumar (2015) only two such studies are present: (1) the study by Newman (2004a) who investigated the application of the Girvan–Newman community detection method on the co-authorship network of the Santa Fe institute, and (2) the study by Pepe and Rodriguez (2010) who investigated mixing patterns in the co-authorship network of the CENS research center. To the best of our knowledge, there are three more studies dealing with the analysis of intra-institutional co-authorship networks: Bellanca (2009), Stefano et al. (2011) and Birnholtz et al. (2013) analyzed co-authorship networks of University of York (UK), University of Salerno (Italy), and two campuses of Cornell University (USA), respectively. Current research information systems (CRISs) offer great opportunities for scientometric studies at the intra-institutional and intra-national levels. For example, Leeuwen et al. (2016) exploited CRIS data to perform bibliographic analysis of research output of a Dutch university. Perc (2010) analyzed the structure and evolution of the co-authorship network encompassing Slovenian researchers which was formed from data stored in SICRIS (a national CRIS system for Slovenia). We are not aware of any study of intrainstitutional research collaboration based on co-authorship networks constructed from CRIS-based data sources. In this paper, we present a hybrid methodology for the analysis of intra-institutional coauthorship networks extracted from CRIS databases. This means that nodes in an analyzed network represent researchers who are institutionally organized into research departments within the corresponding institution. The proposed methodology can be applied to any enriched intra-institutional co-authorship network, i.e. intra-institutional co-authorship networks whose nodes and links are enriched with both domain-independent metrics (metrics used in analysis of complex and social networks which are directly computed from the network) and domain-dependent metrics (e.g. metrics of researcher productivity and competency, collaboration strength and timespan). Using the proposed methodology, we analyze the intra-institutional co-authorship network of the Faculty of Sciences, University of Novi Sad, Serbia (from this point on we use the term ‘‘FS-UNS’’ to denote this particular faculty). The primary goal of the analysis is to investigate how the organizational structure of the institution affects research collaboration and productivity. The network is constructed from the institutional CRIS called CRISUNS. 1 We exploited all benefits that the implementation of CRIS-UNS provides: (1) no name disambiguation problems, and (2) the categorization of publications according to the rule book prescribed by the Serbian Ministry of Science. The rest of the paper is structured as follows. ‘‘Methodology’’ section presents our methodology to analyze intra-institutional research collaboration networks. Researcher evaluation metrics used in our case study are briefly explained in ‘‘Researcher evaluation metrics’’ section. ‘‘Case study’’ section describes the background of our case study. The obtained results are presented and discussed in ‘‘Results and discussion’’ section. Finally, the last section concludes the paper. Methodology Our methodology for the analysis of intra-institutional co-authorship networks is based on the combination of (1) domain-independent metrics and methods used in analysis of complex networks, (2) domain-dependent metrics of researcher productivity and collaboration strength, and (3) non-parametric statistical tests applied to the sets of metric values of independent groups of nodes/links. An intra-institutional co-authorship network is an undirected and weighted graph where link weights express the strength of research collaboration. The following schemes are commonly used to assign weights to co-authorship links: • Straight scheme where two researchers are connected by a link of weight w if they coauthored exactly w different research papers (Batagelj and Cerinsˇek 2013). • The Salton scheme which is a normalized variant of the previous scheme, i.e the weight of a link is in the interval (0, 1] and proportional to the number of joint publications (Lu and Feng 2009), and • The Newman scheme which takes into account the total number of authors in multiauthored publications (Newman 2004b). In our study of the FS-UNS co-authorship network this scheme is used to determine link weights. We assume that nodes in the analyzed co-authorship network are enriched with metrics quantifying productivity, collaboration and institutional importance of corresponding researchers. The same applies for links that are enriched with their importance within the network and timespan (the number of years that passed from the first to the last joint publication of authors connected by the link). Considering the standard organizational structure of research institutions, we assume that each researcher belong to exactly one research department within the institution. Consequently, we distinguish between two types of links: 1. intra-department links co-authorship links connecting researchers belonging to the same research department, and 2. inter-department links co-authorship links connecting researchers from different departments. The methodology proposed in this paper consists of four general steps: 1. The identification and analysis of connected components in the network in order to evaluate the overall collaborative cohesiveness of the institution, 2. The analysis of collaborative cohesiveness of research departments based on graph clustering evaluation metrics, 3. The analysis of inter-department links in order to evaluate research collaboration between departments, as well as to detect differences between researchers involved in inter-department collaborations and researchers whose collaboration is bounded to their own departments, and 4. The comparison of departments relying on researcher evaluation metrics in order to detect similarities and differences in their productivity, collaboration and institutional importance. A connected component of an undirected network is a maximal set of mutually reachable nodes, i.e. there is a path connecting each two nodes in the component. Connected components can be detected using classical graph traversal algorithms (Breadth First Search or Depth First Search). We distinguish between two types of connected components: isolated nodes and non-trivial connected components of research collaboration. Definition 1 (Isolated node) A node in an intra-institutional co-authorship network is called isolated if it is not connected to any other node in the network, i.e. its degree centrality is equal to zero. Definition 2 (Non-trivial component) A connected component in an intra-institutional co-authorship network is considered non-trivial if it encompasses more than one node. Isolated nodes represent researchers who have never collaborated with their institutional colleagues. Therefore, a large number of isolated nodes in the network indicates a poorly connected institutional research community. The size of the largest non-trivial connected component is also an indicator of the overall collaborative cohesiveness of the institution. Definition 3 (Giant component) A non-trivial connected component is considered giant if it encompasses a vast majority of nodes in the network. The existence of giant connected components is one of the main features of co-authorship networks (Newman 2001a, b, c, 2004b; Barabasi et al. 2002; Bettencourt et al. 2009; Perc 2010). The absence of a giant connected component in an intra-institutional research collaboration network implies a poorly cohesive institutional community of researchers, further indicating that the institution is still in an early phase of its scientific development. The structure of connected components can be characterized by various metrics proposed under the framework of complex network theory such as the clustering coefficient, the characteristic path length and the assortativity index (Boccaletti et al. 2006; Savic´ et al. 2014). Another common characteristic of co-authorship networks is the existence of community or cluster structure (Girvan and Newman 2002; Newman 2004a; Leskovec et al. 2009; Savic´ et al. 2015) where clusters are viewed as subsets of nodes that are more densely internally connected than with the rest of the network. In recent times significant research efforts have been devoted to the development of community detection methods (Fortunato 2010; Chen et al. 2015) and graph clustering evaluation (GCE) metrics which quantify the quality of detected communities (Leskovec et al. 2010). For intra-institutional co-authorship networks there is also another notion of communities that is determined by the organizational structure of research institutions. Such institutionally determined research communities are not necessarily strong clusters in co-authorship networks. Namely, if a research department exhibits a low degree of collaborative cohesiveness then the researchers belonging to the department form a poor cluster in the co-authorship network. Therefore, we adopted three GCE metrics to quantify collaborative cohesiveness of research departments: internal density (ID), weighted conductance (WC) and weighted Flake degree fraction (WFDF) (Leskovec et al. 2010). To formally define aforementioned cohesiveness metrics we will assume that • G denotes an intra-institutional co-authorship network, • D an arbitrary research department (a subset of nodes of G), • r an arbitrary researcher within D, and • G[D] the sub-network of G induced by D, i.e. G[D] encompasses researchers from D and all intra-department links between them. The density of a network is the number of links in the network divided by the maximal number of links that the nodes can form. Definition 4 (Internal density) The internal density of research department D is the density of G[D], i.e. ID ðDÞ ¼ 2l=nðn 1Þ; where n and l are the number of nodes and links in G[D], respectively. Higher values of ID indicate more cohesive departments. If ID ðDÞ ¼ 0 then D consists solely of isolated nodes (no research collaboration within D). On the opposite side, ID ðDÞ ¼ 1 implies that D is a clique which means that each two researchers from D co-authored at least one paper together. The strength of research collaboration of researcher r with other researchers within the institution can be estimated by its weighted degree in the co-authorship network. Moreover, we can distinguish between weighted intra-department and weighted inter-department degrees which reflect the total strength of research collaboration of r with researchers from his/her department and researchers from other departments, respectively. Definition 5 (Weighted intra-department degree) The weighted intra-department degree of researcher r, denoted by wintraðrÞ, is the sum of weights of intra-department links incident to r. Definition 6 (Weighted inter-department degree) The weighted inter-department degree of researcher r, denoted by winterðrÞ, is the sum of weights of inter-department links incident to r. Definition 7 (Weighted degree) The weighted degree of researcher r, denoted by w(r), is the sum of weights of all co-authorship links incident to r, i.e. wðrÞ ¼ wintraðrÞ þ winterðrÞ. Definition 8 (Total weight) Let L0 be a subset links in the network. The total weight of L0 is the sum of weights of all links contained in L0. Definition 9 (Weighted conductance) The weighted conductance of D is equal to the total weight of inter-department links incident to nodes in D normalized by the total weighted degree of nodes in D, i.e. WC ðDÞ ¼X r2D winterðrÞX r2D wðrÞ ¼X r2D winterðrÞ X r2D winterðrÞ þX r2D wintraðrÞ : Definition 10 (Weighted Flake degree fraction) The weighted Flake degree fraction of D is the fraction of researchers in D whose weighted intra-department degree is strictly higher than weighted inter-department degree, i.e. WFDF ðDÞ ¼fr : r 2 D ^ wintraðrÞ[winterðrÞg D : The main difference between ID and the other two GCE metrics (WC and WFDF) is that ID takes into account only intra-department links, while WC and WFDF consider both intra-department and inter-department links. The main advantage of WC and WFDF compared to all other known GCE metrics is that they enable the classification of research departments according to the Radicchi notion of communities in complex networks (Radicchi et al. 2004). The Radicchi notion of communities extends the concept of highly cohesive subgraphs known as LS-sets, and nowadays it is one of widely used standards for the evaluation of community detection techniques (Fortunato 2010). Department D can be viewed as a Radicchi strong cluster of researchers if each researcher in D has established stronger collaboration within D than with researchers from other departments, i.e. D is a Radicchi strong cluster ()ð8r 2 DÞ wintraðrÞ[winterðrÞ () WFDF ðDÞ ¼ 1:0 On the other hand, D is a Radicchi weak cluster if the total strength of intra-department collaboration within D is higher than the total strength of collaborations between D and all other departments, i.e. D is a Radicchi weak cluster ()X r2D wintraðrÞ[X r2D winterðrÞ () WC ðDÞ\0:5 Each Radicchi strong cluster is also Radicchi weak, while the converse is not necessarily true. The previously defined GCE metrics will be illustrated on the hypothetical intra-institutional research collaboration network shown in Fig. 1. The network depicts a small research organization that consists of 10 researchers (denoted by letters from A to J) which are institutionally organized into three research departments denoted by P, Q and R. It can be observed that the network contains 9 intra-department and 5 inter-department links. There are 4 links incident to researcher A: 2 intra-department links (A $ B and A $ C) and 2 inter-department links (A $ D and A $ F). The weighted intra-department degree of A is wintraðAÞ ¼ wðA $ BÞ þ wðA $ CÞ ¼ 9 þ 3 ¼ 12, while its weighted inter-department degree is winterðAÞ ¼ wðA $ DÞ þ wðA $ FÞ ¼ 6 þ 2 ¼ 8. Therefore, we can conclude that A has established stronger collaboration with researchers from its own department than with researchers from other departments. Department P is a clique: every two researchers from P have established research collaboration. Therefore, the internal density of P is equal to 1. P is also a Radicchi strong cluster: for each researcher p 2 P we have that wintraðpÞ[winterðpÞ (wintraðAÞ ¼ 12, winterðAÞ ¼ 8, wintraðBÞ ¼ 14, winterðBÞ ¼ 1, wintraðCÞ ¼ 8, winterðCÞ ¼ 0). Department R is not a Radicchi strong cluster: WFDF ðRÞ ¼ 3=4 since G is the only researcher from R whose weighted intra-department degree is lower than the weighted inter-department degree. However, this department is a Raddichi weak cluster: X r2R winterðrÞ ¼ 9;X r2R wintraðrÞ ¼ 20 ) WC ðRÞ ¼ 9=29\0:5 Department Q is neither Raddichi strong nor Radicchi weak: WFDF ðQÞ ¼ 1=3 (researchers D and E have lower weighted intra-department than weighted inter-department degrees) and WC ðQÞ ¼ 18=28[0:5. Our example clearly shows that internal density cannot fully reflect the cohesiveness of departmental research collaboration. Namely, departments Q and R, being drastically different considering their cohesiveness, have the same value of internal density (ID(Q) = ID(R) = 2/3). Relying on internal density we can detect whether researchers from some department form a clique (the most cohesive research collaboration structure) and measure the degree to which organizational units deviate from being cliques, but this metric cannot separate strong from poor research collaboration clusters. On the other hand, WC and WFDF are complementary measures of collaborative cohesiveness: • Using WFDF it can be checked whether research departments possess a strong degree of collaborative cohesiveness (i.e. Raddichi strong clusters), but this measure cannot separate poorly (non Radicchi weak) from moderately cohesive (Radicchi weak) research departments. • WC can indicate poorly cohesive research departments, but it cannot separate strongly from moderately cohesive research departments. To evaluate inter-department collaborations we analyze the structure of a departmental collaboration network, i.e. the network of research collaboration between research departments within the institution. The departmental network is constructed from the coauthorship network by the following rules: 1. The nodes of the departmental network represent research departments of the institution. 2. Two departments A and B are connected by an undirected, weighted link if at least one researcher from A has collaborated with at least one researcher from B. 3. The weight of the link connecting A and B is equal to the sum of weights of all links connecting researchers from A to researchers from B in the co-authorship network. For example, the departmental collaboration network of the institution shown in Fig. 1 consists of three nodes P, Q and R and two links: the link connecting P and Q and the link connecting Q and R. The weight of both links is equal to 9. In our methodological framework, the comparison of two independent groups of nodes/ links in the network is based on the application of the Mann-Whitney U test (Mann and Whitney 1947) and accompanying probabilities of superiority (Erceg-Hurn and Mirosevich 2008). The MWU test belongs to the class of non-parametric statistical procedures which means that it does not assume any particular distribution of compared samples. Let M be an arbitrarily selected node/link metric (a metric of researcher productivity, collaboration or institutional importance in case of nodes; a metric of collaboration strength, timespan or institutional importance in case of links), and let G1 and G2 be the sets of M values for two independent groups of nodes/links. The MWU test is a test of stochastic superiority, and consequently it can be used to check the null hypothesis that the values in G1 do not tend to be systematically greater or smaller than the values in G2. The test is based on the U statistic which is the number of times a value from G2 precedes a value from G1 in the ranked sequence of values from both groups. Under the null hypothesis U closely follows a normal distribution. We use the MWU to examine differences between: • intra-department and inter-department collaboration links, • researchers involved in inter-department collaborations and researchers who do not collaborate with researchers from other departments, and • researchers from two different departments. For the comparison of more than two departments we rely on the Kruskal-Wallis ANOVA test (Kruskal and Wallis 1952) which is a generalization of the MWU for more than two samples. For each conducted MWU test we record two probabilities of superiority to quantify effect size: • PS1 ¼ Pðg1[g2Þ, where g1 and g2 are randomly selected values from G1 and G2, respectively, and Pðg1[g2Þ denotes the probability that g1 is strictly larger than g2, and, • PS2 ¼ Pðg2[g1Þ. Obviously, PS1 þ PS2 ¼ 1 Pðg1 ¼ g2Þ, where Pðg1 ¼ g2Þ denotes the probability that g1 is equal to g2. The probabilities of superiority indicate the degree of stochastic dominance of one group over another and they can be computed in a straightforward manner (by comparing each value from G1 to each value from G2). Researcher evaluation metrics As emphasized in the previous Section, each researcher in a co-authorship network can be characterized by several metrics. Table 1 shows the list of researcher evaluation metrics used in our case study. The first productivity metric shown in Table 1 is based on the evaluation of scientific papers by the rule book prescribed by the Serbian Ministry of Education, Science and Technological Development. The main idea of the rule book is that publication venues indicate scientific importance of publications. The rule book defines several categories of publication venues. Each category corresponds to a certain number of points that are assigned to publications according to their venues. For example, • papers published in the top 30% SCI ranked journals in appropriate scientific discipline are worth 8 points, • papers published in journals that are between the top 30% and top 50% SCI ranked journals are worth 5 points, • papers published in journals with impact factor that are not among the first 50% SCI ranked journals are worth 3 points, while • papers published in proceedings of international conferences give their authors 1 point per paper. The Serbian research competency index for a Serbian researcher is then defined as the sum of points of publications he/she (co)authored. This metric is officially used as one of the criteria in the process of academic promotions at Serbian universities and national research centers, as well as in researcher evaluation within Serbian national research projects. To evaluate the productivity of researchers we also rely on standard productivity measures: normal count, fractional (adjusted) count, and straight count (Lindsey 1980). Normal count gives every author of a publication one point, straight count assigns all the credit to the first author only, while fractional count assigns credit equal to 1/n to each of the n co-authors. To measure the degree of research collaboration at the individual level we use three metrics: 1. Degree centrality (LCOLL) which is a domain-independent local centrality measure equal to the number of links incident to a node in the co-authorship network. This measure is equivalent to the number of local (intra-institutional) co-authors and reflects the degree of researcher’s intra-institutional collaboration. 2. The number of external co-authors (ECOLL) which reflects the degree of researcher’s inter-institutional collaboration. 3:. The total number of co-authors which is the sum of LCOLL and ECOLL. The institutional importance of researchers considering the underlying social structure of intra-institutional research collaboration can be measured by domain-independent global centrality metrics. In this study we rely on the betweenness centrality measure (Freeman 1977; Ko´sa et al. 2015). The main intuition behind the measure is that a node in the network can be considered important if it is located on a large number of shortest paths between randomly selected nodes. This means that nodes having high betweenness are in the position to maintain and control the spread of information across the network. The betweenness centrality of a node z is computed by the following formula BET ðzÞ ¼ X x;y2V;x6¼y6¼z rðx; y; zÞ rðx; yÞ ; where V is the set of nodes in the network, rðx; yÞ is the total number of shortest paths connecting x and y, and rðx; y; zÞ is the total number of shortest paths connecting x and y that pass through z. If a network has a clustered or community organization then nodes with high betweenness tend to be located at the intersections of clusters, which means that they play the role of brokers which connect together various different parts of the network. On the other hand, nodes with low betweenness centrality are typically located on the periphery of the network. Betweenness centrality can also be used to quantify the importance of links in the network. The betweenness centrality of link l is the fraction of shortest paths between randomly selected nodes that pass through l. Case study The Faculty of Sciences in Novi Sad (FS-UNS) is an educational and scientific institution established in 1969. FS-UNS is the second largest of four public faculties of sciences in Serbia, a relatively small European country (approximately 7 million inhabitants) located at the crossroads between Central and Southeast Europe. The faculty consists of five research departments listed in Table 2. The main goal of our case study is to investigate how institutional organization affects intra-institutional research collaboration using the methodology proposed in this paper. It can be observed that the institutional organization of FS-UNS directly corresponds to general scientific disciplines. Therefore, the analysis of research collaboration at FS-UNS can also indicate collaboration patterns characteristic for fundamental scientific fields. The co-authorship network of researchers currently employed at FS-UNS was constructed from the author and publication records contained in the institutional research information system called CRIS-UNS (Ivanovic´ et al. 2010). CRIS-UNS was developed following the recommendations of the non-profit organization euroCRIS.2 Each researcher employed at FS-UNS is institutionally obligated to have his/her CRIS-UNS record that contains all institutionally relevant data (academic rank, research department within the institution, gender, the year of birth, etc.), and periodically update (at least once at the end of a year) his/her references. Therefore, the system practically contains the complete research output of all researchers currently employed at FS-UNS.3 CRIS-UNS is an authorarticle- centered bibliography database. This means that researchers registered in the CRISUNS system have unique identifiers which appear in the CRIS-UNS publication records. When adding a new publication, a researcher has to select co-authors among the researchers registered in the system. Also, he/she can create a profile for an external coauthor (researcher not affiliated to FS-UNS) if the external co-author is not registered in the system. Therefore, each publication recorded by one author is automatically associated to all other co-authors. Each publication record in the CRIS-UNS database contains the following information: an unique publication identifier, the complete list of author identifiers, publication year, title, publication type, and information about publication venue. CRIS-UNS supports the evaluation of publications according to the rule book prescribed by the Serbian Ministry of Education, Science and Technological Development (Ivanovic´ et al. 2011, 2012). These evauations are also stored in publication records and used to compute the Serbian research comptency index for individual researchers. The extraction of the co-authorship network from CRIS-UNS data is a straightforward task since authors are uniquely identified in publication records, and consequently there are no name disambiguation problems. The FS-UNS co-authorship network was constructed from 14986 publications authored by 423 FS-UNS researchers and their 5690 external coauthors. The distribution of researchers per research departments is given in Table 3. To assign link weights we used the weighting scheme proposed by Newman (2004b). Namely, two researchers A and B are connected by a link of weight w computed by the following formula: w ¼X k2J 1 nk 1 ; where J is the set of papers jointly authored by A and B, and nk is the number of authors of paper k. The main property of the Newman scheme is that it does not ignore the total number of authors in multi-authored publications: each joint publication of A and B adds some weight to the overall strength of collaboration between them, but the more authors a joint publication has the less weight is added. Results and discussion The FS-UNS co-authorship network contains 423 nodes (researchers currently employed at FS-UNS) and 2856 links (research collaborations). The researchers from FS-UNS have collaborated with 5690 different researchers not affiliated to FS-UNS with whom they established 34007 research collaborations in total. In this Section we present the results of the analysis of the FS-UNS network following the methodological framework described in ‘‘Methodology’’ section. Connected component analysis Using the BFS graph traversal algorithm we identified connected components in the network. The network consists of 15 connected components where 14 of them are isolated nodes (FS-UNS researchers whose entire production consists of solo-authored publications). This means that the network possesses a giant connected component encompassing a strong majority of FS-UNS employees (96.69% of them). The existence of the giant component implies that 1. the FS-UNS researchers form a strongly cohesive institutional community of researchers, and, 2. the scientific output of the institution is not a product of many research groups that do not collaborate among themselves. The giant connected component exhibits small shortest path lengths (the average distance between two randomly selected nodes is 3.32), and a drastically larger clustering coefficient than the clustering coefficient of comparable Erd}os-Renyi random graphs. The clustering coefficient of a random graph with N ¼ 409 nodes and L ¼ 2856 links is equal to 2L=NðN 1Þ ¼ 0:034, while the clustering coefficient of the giant connected component of the FS-UNS network is equal to 0.566. Therefore, we can conclude that the FS-UNS network exhibits the Watts-Strogatz small-world property (Watts and Strogatz 1998; Newman 2001b). The Newman index of degree assortativity (Newman 2002) is equal to 0.24 implying that highly intra-collaborative researchers moderately tend to be directly connected among themselves. Cohesiveness of research departments The FS-UNS network can be partitioned according to the organizational structure of the faculty into sub-networks that contain researchers belonging to the same research department. By investigating characteristics of these sub-networks we can derive conclusions related to the collaborative cohesiveness of research departments. Therefore, for each departmental sub-network we performed connected component analysis and computed graph clustering evaluation metrics quantifying group cohesion. The results are summarized in Table 4. We can see that the following holds for each of the departments: 1. The corresponding departmental sub-network has a giant connected component encompassing all researchers from the department except a small number of isolated nodes. The FS-UNS co-authorship network contains 14 isolated nodes, while there are 19 isolated nodes in the departmental sub-networks. Therefore, 5 FS-UNS researchers (1.22%) involved in inter-department collaborations have never collaborated with researchers from their own departments. 2. The total number of intra-department links is higher than the total number of interdepartment links. 3. The total weight of intra-department links is drastically higher than the total weight of inter-department links implying that the departments are Radicchi weak communities (moderately strong clusters in the network). In other words, for each department we have that the total strength of research collaboration within the department is higher than the total strength of research collaborations with other departments. The obtained values of the weighted Flake degree fraction indicate that all departments are very close to being Radicchi strong communities. Namely, a large majority of FS-UNS researchers have established stronger collaborations within their own departments than with researchers from other departments. However, in each department there is a small number of researchers (from 1 to 4) which have collaborated stronger with colleagues from other departments than with colleagues from the department they institutionally belong. In the case of institutions such as faculty of sciences, the identification of such ’’out-ofgroup’’ researchers is especially important since they have proven that they possess skills and knowledge highly valuable for interdisciplinary research. For example, four ’’out-ofgroup’’ researchers from the department of mathematics are experts in statistical modeling, discrete mathematics and partial differential equations applicable in biology, chemistry and physics. From the data presented in Table 4 we can also observe differences between the departments. Namely, we can see that the department of geography is the most cohesive and the most closed research division since it has the highest internal density of intradepartment links (and consequently the highest average internal degree) and at the same time the lowest weighted conductance. In other words, researchers from this department tend mostly to collaborate among themselves, while collaborations with researchers from other departments are relatively rare compared to other FS-UNS departments. For example, the department of physics possessing a smaller number of researchers has a drastically higher (almost two times higher) degree of inter-department collaboration. The department of mathematics and informatics possesses the weakest degree of internal cohesion. This fact can be explained by a global trend of relatively low collaboration in mathematics compared to other natural and experimental sciences (Grossman 2002; Barabasi et al. 2002). The strongest intra-department collaboration can be observed for the department of physics and the department of chemistry whose average internal weighted degrees are significantly higher compared to other departments. Inter-department collaborations The FS-UNS network contains 582 inter-department links (20.4% of the total number of links in the network). Table 5 shows the results of the Mann-Whitney U test for the differences between intra-department and inter-department links. It can be seen that: 1. Intra-department collaborations have significantly higher weight and timespan compared to inter-department collaborations. This means that intra-department collaborations tend to be significantly stronger and longer in time than interdepartment collaborations. 2. On the other hand, inter-department collaborations tend to have significantly higher betweenness centrality implying that they are more important for the overall cohesiveness and connectedness of the underlying social structure of the institution. Further, we divided the nodes of the network into two groups: 1. G1 that contains researchers involved in inter-department collaborations. A researcher belong to this category if he/she co-authored at least one paper with a researcher belonging to different department, and 2. G2 that contains researchers who have never collaborated with colleagues from other departments. G1 encompasses 227 FS-UNS researchers (53.7%), while the rest of them belong to G2. We applied the Mann-Whitney U test in order to identify differences between those two groups of researchers. The results are summarized in Table 6. It can be seen that the null hypothesis of the test is rejected for each considered metric. Moreover, researchers belonging to G1 tend to have drastically (Avg(G1) Avg(G2)) and systematically (PS1 PS2) higher values of all considered researcher evaluation metrics. Therefore, we can conclude that researchers involved in inter-department collaborations tend to be drastically more productive (by all considered productivity metrics), collaborative (both locally and externally) and institutionally important than researchers who only collaborate with colleagues from their own departments. Since the FS-UNS network exhibits an assortative degree mixing pattern and the researchers from G1 tend to have systematically higher degree centrality (the LCOLL metric in Table 6), we can also conclude that the FSUNS network has a core-periphery structure where the researchers from G1 constitute the core of the network. The departmental collaboration network of FS-UNS is shown in Fig. 2. The widths of links in the figure are proportional to their weights, while the numeric link labels show the number of links between researchers in the FS-UNS co-authorship network. It can be seen that the departmental collaboration network is a fully connected graph which means that each department maintains research collaboration with all other departments. However, the strength of departmental collaboration links is highly unbalanced. It can be observed that the strongest departmental collaboration is between the department of chemistry and the department of biology—150 different co-authorship links connect researchers from those two departments. The total strength of collaboration between those two departments is equal to 13,256 and it is more than four times higher than the second most strongest departmental collaboration (the collaboration between the department of chemistry and the department of physics). The weakest strength of collaboration is between the department of chemistry and the department of geography (5 links in the FS-UNS co-authorship network whose total weight is 1.09). It can be noticed from Fig. 2 that several departmental collaborations have very weak intensity suggesting that there is a lot space for the improvement of inter-departmental cooperation at FS-UNS. Evaluation of researchers For each of researcher evaluation metrics listed in Table 1 we computed the distribution of metric values considering all nodes in the FS-UNS network. The summary of descriptive statistics of the distributions is given in Table 7. It can be seen that all distributions are highly skewed to the right (skewness[1). The high skewness of the distributions implies that there are deep inequalities in productivity, collaborative behavior and institutional importance of FS-UNS researchers. In other words, there are FS-UNS researchers whose degrees of productivity, collaboration and institutional importance are much higher than the average values. For example, the productivity of FS-UNS researchers measured by the normal counting scheme follows the famous Pareto principle: 80% of the complete FSUNS scientific production is authored by 20% of FS-UNS researchers (see Fig. 3). The differences in productivity, collaboration and institutional importance between researchers from different departments are analyzed using the Kruskal-Wallis ANOVA test. The results are summarized in Table 8 which also shows the average value of each metric for each FS-UNS department. From the presented data we can infer the following: • There are statistically significant differences between researchers from different departments considering their productivity estimated by the Serbian research competency index and the normal counting scheme, but significant differences are absent when fractional and straight counting schemes are used to measure research productivity. The highest average productivity measured by SRCI and PRO has the department of physics, while the lowest values are associated with the department of mathematics and informatics. The absence of statistically significant differences in productivity measured by fractional and straight counting schemes between researchers from different departments indicates that SRCI and PRON are biased measures of research productivity. • There are statistically significant differences in the degree of both intra-institutional and inter-institutional collaboration of researchers from different departments. The highest average degree of intra-institutional research collaboration exhibits the department of geography, while the highest average degree of inter-institutional collaboration is associated with the department of chemistry. • There are not statistically significant differences in the institutional importance of researchers from different departments when the institutional importance is estimated by the betweenness centrality metric. In order to detect which departments are stochastically superior considering SRCI, PRON, LCOLL and ECOLL metrics we performed a series of pairwise department comparisons relying on the Mann-Whitney U test. The results are summarized in Table 9 which shows pairs of departments for which the null hypothesis of the MWU test is rejected. We can observe that researchers from the department of physics and department of chemistry have systematically higher values of the Serbian research competency index than researchers from all other departments. Namely, the probability that a randomly selected FS-UNS physicist or chemist has strictly higher SRCI than a randomly selected FS-UNS mathematician, biologist or geographer is almost two times higher than the inverse probability of superiority. Regarding productivity estimated by the normal counting scheme, it can be seen that (1) FS-UNS mathematicians have significantly lower productivity compared to FS-UNS physicists and FS-UNS chemists, and (2) FS-UNS chemists have significantly higher productivity than FS-UNS biologists and geographers. Since there are no statistically significant differences between FS-UNS departments considering productivity estimated using fractional and straight counting schemes, we can conclude that SRCI and PRON are biased towards FS-UNS physicists and chemists. Therefore, those two productivity metrics cannot be used to compare productivity of FSUNS researchers from different departments. This is especially important since SRCI is officially used in the process of academic promotion, but also in many official, administrative decisions (e.g. SRCI is used by the UNS housing commission as one of factors when forming priority lists for flat allocation, and by the Ministry of Science when approving and evaluating national research projects). From the data presented in Table 9, it is evident that the organizational structure of FSUNS has a strong impact on both intra-institutional and inter-institutional research collaboration. We can see that the researchers from the department of mathematics and informatics have drastically lower degree of both intra-institutional and inter-institutional collaboration compared to researchers from other departments. On the other hand, the researchers from the department of geography and the department of chemistry have significantly higher degree of intra-institutional collaboration than researchers from other departments, and at the same time statistically significant differences in the degree of interinstitutional collaboration are absent. Therefore, we can conclude that those two departments actively stimulate intra-institutional research collaboration. Conclusions Current research information systems (CRISs) offer great possibilities to analyze research output and collaboration at the intra-institutional level. However, these possibilities have been rarely exploited in contemporary scientometrics studies. We proposed the hybrid methodology to analyze enriched intra-institutional co-authorship networks constructed from CRIS databases. Our methodology is based on: • connected components analysis and graph clustering evaluation metrics in order to investigate the cohesiveness of intra-institutional research collaboration at the level of the whole institution and the level of departments within the institution, • the structural analysis of departmental collaboration networks induced from intrainstitutional co-authorship networks in order to reveal patterns of collaboration between research departments, and • non-parametric statistical procedures applied to independent groups of nodes/links in the network in order to (1) evaluate the differences between researchers from different departments with respect to their productivity, collaborative behavior and institutional importance, and (2) determine characteristics of researchers involved in interdepartment collaborations. Using the proposed methodology, we analyzed the intra-institutional co-authorship network extracted from the institutional CRIS system of the Faculty of Sciences, University of Novi Sad (FS-UNS). The connected component analysis revealed that FS-UNS is an institutionally cohesive research community with a very small fraction of researchers whose research output entirely consists of single authored publications. The analysis of departmental cohesion based on graph clustering evaluation metrics indicated that a very large majority of FS-UNS researchers have established stronger collaborations within their own departments than with researchers who institutionally belong to other departments. However, researchers involved in inter-department collaborations tend to be drastically more productive, both locally and externally collaborative, and institutionally important compared to colleagues whose collaboration is bounded to their own research departments. Empirically observed distributions of researcher evaluation metrics are highly skewed to the right implying that there are deep inequalities regarding the productivity, collaboration and institutional importance of FS-UNS researchers. Further analysis based on nonparametric statistical tests revealed that there are statistically significant differences in the productivity and collaboration of researchers from different departments, but not regarding their institutional importance. Two FS-UNS departments have significantly higher degree of intra-institutional research collaboration compared to the rest of FS-UNS departments indicating that those two departments actively stimulate intra-institutional research collaboration. Finally, we observed that statistically significant differences in the productivity of researchers from different departments are present when research productivity is estimated using the normal counting scheme and the Serbian research competency index, but absent when research productivity is measured by the fractional or straight counting scheme. Therefore, we can conclude that the Serbian research competency index is not an adequate measure to compare researchers from different scientific disciplines, and it should be avoided in strategic and administrative decision making.

Seeking to ensure that students have access to Bhigh impact^ learning experiences (Kuh 2008), institutions of higher education have aspired to increase the availability and visibility of programs that engage undergraduates in research, scholarship, and creative work (URSCW) (Malachowski et al. 2015). While the growing body of research provides strong evidence that students benefit from participation in such activities (Ishiyama 2002; Jenkins and Healey 2010), it is not clear whether or not institutional aspirations to promote URSCW are matched by institutional support of faculty engagement. Faculty members who serve as mentors in this high-impact practice are the facilitators of undergraduate research, thus making it critical to understand the messaging and corresponding resources and support that faculty members receive about the value of their participation. Thus, the research question for our study was as follows. Do institutions’ public messages about their commitment to undergraduate research, scholarship, and creative work align in formal evaluation and reward systems with the resources available to faculty members and the messages faculty members receive about the value of their work as mentors? This research extends our prior efforts (Baker et al. 2015), which focused on faculty mentors’ experiences, perceptions, and understanding of URSCW. In the study we report here we turned our attention to an examination of the connection between institutional aspirations and the policies and practices related to excellence in undergraduate research, scholarship, and creative work. Relying on policy analysis as a framework, we reviewed the websites of 100 institutions randomly selected from the membership list of the Council of Undergraduate Research (CUR) in order to examine public statements of support for this work. We then compared these public statements to guidelines from CUR for recognizing faculty work in this high impact practice. Mentoring Undergraduates: Realities and Representations The research presented here is situated at the intersection of several bodies of scholarship. We first discuss the critical role of the faculty mentor in undergraduate research, scholarship, and creative work and the ways in which institutions both enable and constrain faculty members as they take on this work. Next, we discuss how colleges and universities communicate with internal and external constituencies via institutional websites. We conclude with a discussion of the connection between promotion and tenure criteria and undergraduate research, scholarship, and creative work. Faculty Mentors and Institutional Constraints A mentoring relationship is one of the key characteristics that differentiates undergraduate research from other educational experiences, and the benefits to students who participate in URSCWare well documented (Dolan and Johnson 2010; Elder and Trapp 2010). Despite the time-intensive nature associated with effective mentorships, Johnson (2007) found that faculty mentors experience positive outcomes including personal satisfaction, fulfillment, and reputational gains for talent development. Other studies showed, however, that mentoring comes at a cost to faculty members. Schwartz (2012) studied pairs of faculty mentors/students in STEM fields and determined that faculty members struggled with the affective and monetary impact of their work. Buddie and Collins (2011) found that, while many of the best practices in the literature address the issues that might improve the undergraduate research experience for students, more institutional support is needed to help faculty mentors manage the associated challenges such as time and funding. In perhaps the largest study of faculty mentors of undergraduate researchers to date, Eagan et al. (2011) analyzed data from nearly 5,000 science faculty members to determine how institutional contexts and individual factors influenced faculty decisions to serve as mentors. They concluded that Bwithout tangible incentives to create research opportunities many faculty [sic] may decide to involve undergraduate students in research projects solely as a result of good organizational citizenship behavior^ (p. 173). The pool of faculty members intrinsically motivated to take on the additional work of mentoring undergraduate researchers, scholars, and artists is limited; and, as institutions seek to develop and sustain more robust undergraduate research programs, campus leaders will need to generate wide support among faculty members by linking incentives and behaviors. Given the importance of institutional context and associated supports, it is possible that institutional differences such as resources and mission influence the infrastructures that support this high impact practice. Structures such as undergraduate research offices and incentives may be more prevalent in baccalaureate institutions, given their focus on undergraduate teaching. However, the converse may be true – that public research institutions are more likely to have the financial resources and organizational capacity to support such structures and rewards, thus making the presence of undergraduate research offices greater at this institutional type.We know of no research that has considered the role of institutional type in shaping internal and external messaging about faculty mentorship of undergraduate research, scholarship, and creative work. Institutional Websites and Communication Networks The earliest studies of college and university websites date to the 1990s, and they focused on the functionality and design of such digital spaces (Hossler 1999; Williams 2000). Poock and Lefond’s study (2001) of how prospective students used websites in the process of selecting and applying to post-secondary institutions underscores the importance of content related to the admissions process and the overall environment of the school, the architecture of the website and its ease of navigation, and a focus on the users’ interests. As digital technologies expand, research continues to emerge about how post-secondary institutions may build relationships with prospective students (see for example, McAllister 2012). Taking a more expansive approach, Middleton et al. (1999) urged institutional leaders to recognize the importance of using websites to communicate not only with prospective students and other external constituencies, but also internal audiences including the faculty and staff. In her study of 40 college and university websites, Meyer (2008a) traced how websites address multiple audiences, including current and prospective students, faculty and staff, alumni, potential donors, and others. She characterized the website of a college or university as the Bvirtual face^ that it chooses to present to Bits virtual visitors, which makes it an important window into the institution, a clue to its priorities, and evidence of how it wishes to be seen^ (2008b, p. 178). Research by Saichaie and Morphew (2014) suggested that institutional websites may represent complex institutional compromises as they seek to present the most appealing image possible. These studies provide a foundation on which to understand the importance of and connection between institutional websites and aspirations as communicated through internal and external messaging. While heeding the cautionary notes of Wilson and Meyer (2009) and Saichaie and Morphew (2014) that websites may not fully represent the reality of the faculty experience, we observe that they are perhaps the single greatest source of communication about an institution’s priorities and programming to diverse stakeholders. Promotion and Tenure and Undergraduate Research Ernest Boyer’s 1990 Carnegie Report, Scholarship reconsidered: Priorities of the professoriate initiated scholarly conversations about faculty roles, responsibilities, and rewards. Scholarship reconsidered urged postsecondary institutions to re-conceptualize the evaluation of faculty members and promotion and tenure processes around four critical domains: discovery, integration, application, and teaching. Several researchers have documented the impact of Scholarship reconsidered and assessed its implementation in a variety of institutional contexts (Griffin 2012; Hutchings et al. 2011; O’Meara 2005). In general, these studies point toward the importance of administrative leadership as institutions undertake the challenging work of revising evaluation policies. The rewards of such efforts, however, can be significant; O’Meara noted that the revision of these policies resulted in greater attention to undergraduate learning. Yet, there is little research that delves into how the work of mentoring undergraduate researchers, scholars, and artists is addressed in promotion and tenure policies. In 2011, the CUR Quarterly, published by the Council for Undergraduate Research, focused on undergraduate research, scholarship, and creative work and promotion and tenure processes. Authors from a range of institutions and disciplines shared accounts of how the work of mentoring undergraduate researchers enters into annual evaluation of faculty. The work of mentoring undergraduate researchers can confound the traditional division of faculty responsibilities into research, teaching, and service (Ronnenberg and Sadowski 2011; Vaughan 2011), thus highlighting several critical issues including how to count a co-authored, peerreviewed publication with an undergraduate researcher or a secured grant to cover the expenses of undergraduate researchers. The contributors raised questions about disciplinary differences in the work of mentoring undergraduate researchers, which can affect how such work is then factored into promotion and tenure decisions (Vaughan 2011). Institutional differences, as well as disciplinary differences, may also affect such recognition. Undergraduate research, scholarship, and creative work are very important activities that support the learning and advancement of undergraduate students. Through membership in CUR and public messages via institutional websites, administrators and campus leaders communicate their commitment to encouraging and supporting student engagement. Yet, much less attention is given to how such aspirations compare with the resources and recognition provided to the faculty mentors who facilitate these kinds of experiences for undergraduates, particularly with regard to faculty evaluation procedures. We now provide an overview of the conceptual framework that guided our research – policy analysis. Conceptual Framework: Policy Analysis and Policy as Networked Discourse There is a strong tradition of policy analysis within colleges and universities as a method of comparing institutional claims to the realities of institutional practices. Allen et al. (2010) gave considerable attention to the potential of critical examinations of policy as a strategy for identifying assumptions and shaping practices. Fugazzotto’s (2009) analysis of the relationship between mission statements as abstract indicators of institutional culture and space allocations as structural sources of evidence about institutional values called to light administrative opportunities to improve the alignment of both cultural and structural expressions of the institutional mission. Similarly, Morphew and Hartley (2006) conducted a thematic analysis of nearly 300 institutional mission statements as public indicators of their values and purposes. More specifically, there is precedent for conducting policy analysis about faculty experiences and outcomes related to the promotion and tenure process. The American Council on Education (Helms 2015) recently released its assessment of promotion and tenure guidelines in regard to the inclusion (or lack of inclusion) of globally focused evaluation criteria, despite the fact that institutional mission statements increasingly emphasize global or international education. We know of no studies, however, that have explored the alignment between public statements of value for undergraduate research, scholarship, and creative work and policies for evaluating faculty members’ engagement. Our focus in this continued line of inquiry (Baker et al. 2015) turned towards understanding if and to what degree a disconnect exists between messages about the importance of URSCW and available resources and incentives, particularly for faculty members.We hypothesized that there would be dissonance in aspirations versus actual practice related to undergraduate research, scholarship, and creative work. Such dissonance, for example, could be the result of institutional pressure from campus leaders and administrators to offer undergraduate research experiences given the known benefits to students, despite the lack of financial resources to support faculty members who engage in such programming. Or it could be the case that programming is created in order to offer a variety of experiential offerings, but policy fails to keep up with actual practice (see Table 1 for a list of Hypotheses). A college or university website is a rich, multi-layered, multi-modal text that communicates vital information to prospective and current students, faculty, staff, and other stakeholders. Through the policy decisions an institution makes about website content, it is also constructing a particular campus reality and particular identities for campus constituents. We thus offer not only a quantitative analysis of how a sample of postsecondary institutions represents the work of faculty mentors, but also consider how the policies embodied in evaluative documents define the role of the faculty involved in this activity. The Study We followed Iverson’s (2012) example of discursive policy analysis to explore normative frameworks, assumptions in language, and limits to policy effectiveness. For our purposes, institutional web pages about undergraduate research and corresponding offices and statements about recognizing and rewarding faculty work in evaluation guidelines comprised the discourse of focus for this analysis. Referring not to websites and faculty handbooks but to the similar document of a university student handbook, Iverson (2012) explained, B[It] is at once a written document that reflects a given reality, an archival ‘snapshot’ yet it also contributes to producing a given campus reality…^ (p. 154). Krippendorff (2013) defined content analysis as Ba research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use.^ (p. 24). He advocated for the technique as an unobtrusive research strategy that allows for the handling of potentially large volumes of unstructured content as data in ways that are context-sensitive. He also emphasized the value of content analysis not only to provide new insights about a topic but also to inform related practice and action. It is a method that can be useful in documenting and making meaning of potential differences (Krippendorff 2013) – in this instance, the differences between public institutional messages about the value of faculty engagement in undergraduate research and those communicated directly to the faculty through evaluation documents. Sample and Selection Criteria We randomly selected 100 colleges and universities (50 baccalaureate institutions, 50 public research universities) from the 2015 Council of Undergraduate Research (CUR) membership list (N = 546). Data collection and analysis occurred between July and November, 2015. These schools, by virtue of their membership, have indicated their support of CUR’s goal and mission of promoting high-quality undergraduate studentfaculty relationships around collaborative research and scholarship. We copied and pasted the member institutions in an excel spreadsheet and numbered each school 1– 546. The second author generated random numbers using www.random.org. The colleges and universities matching the random numbers were selected until we identified 50 public research institutions and 50 baccalaureate institutions. We relied on the Carnegie Classification system to confirm institutional categorizations. Tables 2 and 3 lists the institutions included in this analysis. Data Collection We sought to determine the extent to which institutions’ internal and external messages about their commitment to undergraduate research, scholarship, and creative work align with policy and practice related to faculty evaluation and recognition for mentoring. We applied Krippendorff’s (2013) content analysis process, which includes unitizing, sampling, recording/coding, reducing, inferring, and narrating. We first unitized the information to be treated as data (i.e., the collection and organization of information as collected from websites) including references to faculty engagement in undergraduate research in (a) promotion and tenure documents and (b) top search results within institutional websites for the term Bundergraduate research.^ We then identified our sampling technique, as described above, and limited the sample to 100 randomly selected institutions from the Council for Undergraduate Research membership list. Next, we recorded and coded the data as identified in the unitizing stage into the database for each institution included in the sample. Through independent analysis, interrater reliability, checks and research team meetings, we developed and refined our data reduction strategy and applied that strategy to populating the database. Finally, the completion of the database allowed us to infer meaning from our results and narrate those inferences as responses to our guiding research questions. Data Sources and Variables We relied on CUR’s (2012) Characteristics of Excellence in Undergraduate Research (COEUR) to identify best practices that support and sustain highly effective undergraduate research to guide our variable selection. The 9 variables included in our study are as follows: UR Office According to COEUR (2012), the most highly successful undergraduate research programs are associated with a central office of undergraduate research which serves as the clearing house for campus-wide undergraduate research activities. We identified such a presence by finding an actual office/campus address and a full time director of undergraduate research. Top 3 Link Search We searched main page institutional web pages for Bundergraduate research^ and recorded the top 3 Bhits^ resulting from that search. Given the lack of precedent or resources for best practices in web page content analysis, we determined that publicly available institutional information about undergraduate research would be likely to appear in the top 3 search results for a given search term. Faculty Incentives We searched the undergraduate research office main page for mention of faculty incentives. According to COEUR (2012), such support should include faculty startup funding that Bis commensurate with institutional expectations^ (p. 11) to support scholarly activity and engagement with undergraduate research. We classified and coded faculty incentives in one of two ways: monetary (i.e., startup funds, salary support) and time (i.e., course release). Student Incentives We searched the undergraduate research office main page for details about student incentives. Funds to support dissemination are critical to a robust undergraduate research enterprise, and those institutions which do provide such support (to at least one conference per year) are categorized as having exemplary undergraduate research programs (COEUR 2012). We classified and coded student incentives in one of three ways: monetary (i.e., salary), housing (i.e., summer on campus housing), and dissemination (i.e., funds to support conference travel). Faculty Resources We searched undergraduate research office main pages for links to faculty resources given that such supports are critical to encouraging faculty engagement in undergraduate research (COEUR, 2012). We classified and coded faculty resources in one of two ways: mentoring (i.e., mentoring guides) and student supports (i.e., how to engage undergraduates in research). Student Resources We searched undergraduate research office main pages for links to student resources. We classified and coded student resources in one of three ways: writing (i.e., writing guide, proposal preparation), research (i.e., how to develop a research question), and presentation/dissemination (i.e., how to present your research). Mentoring Recognition We searched the URO Main page and main institutional web page for BMentoring Award^ directly connected to undergraduate research experiences. We also included two broad categories of variables for comparative purposes. Institutional Type Based on findings of undergraduate research experiences, institutional context is an important factor (Eagan et al. 2011). We therefore included public research universities and baccalaureate institutions for comparative purposes in our sample. Enrollment Enrollment data were collected from the Carnegie Classification of Institutions of Higher Education Website (http://carnegieclassifications.iu.edu). Data Analysis We reviewed faculty handbooks, institutional bylaws, evaluation guidelines and criteria, and any related documents available online which outlined the faculty evaluation process and criteria at the institutional and divisional levels. To ensure consistency, we relied on the institutional level evaluation and promotion and tenure criteria for the purposes of our analysis. We searched for the explicit mention of Bmentoring undergraduate research^ in the evaluation criteria. The data were analyzed in the R statistical package (R Core Team 2015). Frequencies and Chi-squared equations were calculated to answer the research questions and test hypotheses. Trustworthiness, Validity, and Reliability To ensure trustworthiness, we relied on three techniques. First, the two lead authors coded two institutions (one public research university, one baccalaureate) separately. After recording and coding the variables listed above for the institutional websites, both authors discussed their codes and developed a training guide for the remaining authors. In-person training occurred with the entire team of authors simultaneously. As a team, we coded an institution together to ask questions and provide further clarification before coding remaining schools separately. Each author coded 20 institutions. Second, we randomly selected 10 % (n = 10) of the institutions to be coded by a second rater. Interrater correlation was calculated to determine agreement. Seven variables had agreement at the level of .82 or higher. Student resources from a undergraduate research office had a .62 agreement. We determined there was an error in the coding for student resources and recoded that variable for all institutions to ensure that student resources were recorded only when they were provided by a undergraduate research office. The final correlations were between .82 and .91 for all variables, which indicated acceptable agreement (Cohen and Cohen 1983). Third, given the pace at which institutional websites change, we took screen shots of each step to ensure consistency and accuracy between the time of data collection and future analyses. Screen shots also aided training as authors talked through interpretation and coding categories. All screen shots were saved in a shared Dropbox folder. Limitations As with any research, there are limitations worth noting. Given our desire to be as consistent as possible, we reviewed websites for explicit language only using a few key search terms such as Bundergraduate research^ when searching for undergraduate research offices or Bmentoring undergraduate research^ when reviewing evaluation criteria. It is possible that such search terms did not redirect us to the variables of interest at the institutional level given that undergraduate research offices or evaluation criteria may be present on a divisional basis or undergraduate research may be decentralized at the departmental or even programmatic levels. We also note that faculty resources to support mentoring undergraduates, for example, may be located on an institutional resources page found under Academic Affairs, rather than on a website dedicated to undergraduate research. Lastly, we note generalizability may be an issue with other institutional types, particularly non-CUR members. Findings Our main aim in conducting this research was to examine the alignment between public statements of value for undergraduate research, scholarship, and creative work as communicated through institutional websites and policies for evaluating faculty members’ engagement in such work through P&T criteria. Undergraduate Research Office Steps 1 and 2 of analysis focused on whether URSCW appeared in the top 3 links when we searched for Bundergraduate research^ on the main institutional web page. When reviewing such websites, 39 of the 100 institutions had an undergraduate research or undergraduate research office link appear in the top three hits. There was a significant difference by institutional type in favor of public research institutions χ2 = 20.35 (df= 1, p = 6.467e-06) being more likely to display such a hit. As part of Step 3, we checked for the presence of an actual undergraduate research office (versus an undergraduate research web page). Forty-three institutions had an undergraduate research office, while 57 did not. We calculated a chi-square equation to determine if there was a relationship between having an undergraduate research office with institutional type (H1a). Public universities were significantly more likely to have an undergraduate research office compared to the baccalaureate institutions in our sample χ2 = 19.75 (df = 1, p = 8.84e-06). Undergraduate enrollment was coded into ten categories based on the natural break in size. The categories were less than 1,000; 1,001–2,000; 2,001–3,000; 3,001–6,000; 6,001–10,000; 10,001–15,000; 15,001–25,000; 25,001–30,000; 30,001–40,000; and greater than 40,000. Chi-square calculations revealed that enrollment size was significantly related to having a URO (H1b), and this relationship was significant χ2 = 25.746, (df = 9, p = 0.002248). Post-hoc analysis revealed institutions with an enrollment from 10,000 to 30,000 were more likely than those institutions with lower (or greater) enrollments to have an undergraduate research office and were responsible for the significant finding. Support for Students and Faculty Institutions with larger student enrollments, as compared to those institutions with lower student enrollments, may have more resources to provide student and faculty incentives to participate in undergraduate research. However, Chi-squared analyses showed no significant difference between the institution’s enrollment and undergraduate research office incentives or resources for students or faculty. Between 75 and 95 % of the URO offices provided student incentives (N = 33) such as grants and scholarships and other resources (N = 41) including workshops on proposal preparation, writing guides, and presentations guides (see Fig. 1). Less attention was given to faculty members with just over 25 % of undergraduate research offices providing incentives for faculty (N = 12) and only about 50 % of undergraduate research offices providing faculty resources (N = 23), or awards regardless of institution enrollment (H2a/H2b and H3a/H3b). Thus, the hypotheses about institutional size and concomitant resources, incentives, and supports were not supported. Recognition in Promotion and Tenure Our analysis of promotion and tenure guidelines revealed that only 14 institutions explicitly mentioned Bmentoring undergraduate research^ in any evaluation criteria regardless of institutional type (see Table 4), and there was no statistical difference by institution type (H4). When Bmentoring undergraduate research^ was listed in promotion and tenure documents, it was most often categorized under teaching. We conducted a post-hoc analysis to determine if institutions with undergraduate research offices were more likely to explicitly mention Bmentoring undergraduate research^ in their promotion and tenure documents, and the answer was no. Finally, we noted language in promotion and tenure evaluation criteria that mentioned mentoring, though not explicitly in the context of undergraduate research. Twenty of the 100 institutions in our sample included some mention of mentoring (see Table 4). Sample statements included BTeaching includes supervising directed inquiries, honors projects and/ or other forms of student research or creative activity,^ BTeaching: one-on-one in research, independent study, cooperative education, internships and practica,^ and BTeaching is not limited to classroom instruction, but includes activities such as supervising, mentoring, and advising students.^ All mentions of mentoring, outside of the context of undergraduate research, were listed under the BTeaching^ category in promotion and tenure criteria. Discussion Undergraduate research has been increasingly embraced by colleges and universities as a high impact practice; yet it requires a significant investment on behalf of the institution as campus leaders and administrators create the necessary infrastructures and incentives to support such activities. However, we find that the reality of the necessary support has yet to catch up with institutional aspirations, thus resulting in different audiences receiving different messages – the increased offerings of undergraduate research opportunities and the associated supports and incentives for students far outweigh those made available to faculty members as evidenced by information communicated through institutional websites and evaluation policies and criteria. Additionally, our findings reveal that, while the institutions in our study have aspirations to promote and engage undergraduates in research, this aspiration has yet to translate fully into creating the infrastructures and subsequent changes to evaluation policies needed to support the faculty members who engage in these activities.We agree with O’Meara et al. (2015), who noted that the promotion and tenure process, as a part of the larger reward system present in higher education institutions, reflects institutional values, aspirations, privileges, and power structures. What we found, at the time of this study, in the promotion and tenure policies related to undergraduate research, scholarship, and creative work, has failed to keep up with practice, at least for the institutions included in our study. Implications for Research We believe that the findings from our study contribute to research about undergraduate research, scholarship, and creative work and the use of institutional websites as an important window into understanding aspirations versus actual practice. First, our study is the first of which we are aware that has investigated how institutions communicate with faculty about the importance of URSCW, how institutional policies support or impede faculty involvement in such practices, and how participation is factored into faculty evaluation systems. Consistent with the work of Schwartz (2012) and Buddie and Collins (2011), our findings indicate that faculty who participate in undergraduate research, scholarship, and creative work may do so despite a lack of institutional support.We agree, however, with Eagan et al. (2011) that change needs to occur as campus leaders increase their expectations about the types and amount of undergraduate research experiences made available to students, which ultimately impacts academic work and faculty responsibilities. Successful mentoring relationships require training, support, and recognition for both individuals involved. We wonder if more faculty members might engage more students if there were institutional support such as a centralized Office of Undergraduate Research that provides programming or resources to help faculty develop the intellectual and affective competencies to serve as mentors as well as material incentives such as stipends and recognition through faculty evaluation policies. Second, our findings revealed that the public research institutions in our sample were more likely to have undergraduate research office and undergraduate research appear in the top three hits when searching for that term. We found this finding interesting given that the emphasis of teaching and working with undergraduates is perceived to be more predominant at baccalaureate institutions. Perhaps baccalaureate institutions believe they have such experiences infused throughout their educational offerings, thus removing the need for undergraduate research offices at this institutional type (Turner et al. 2008). Within the public research universities included in our sample, it appeared that the really large institutions (enrollment over 30,000) also did not have such structures. It might be the case that these institutions, given their size, decentralize undergraduate research leaving such experiences within honors colleges or academic departments. Third, a strength of our study is that we relied on what institutions report they do rather than surveying individuals. Our findings support and extend research that relies on institutional websites as useful sources of data in that choices about what is (and is not) included provides a telling story to internal and external constituencies (Taylor and Morphew 2010). Our results align with the work of Saichaie and Morphew (2014) and Middleton et al. (1999) who showed that institutional websites convey important messages and an institutional image to external and internal audiences. Furthermore, they serve as the Bvirtual face^ as colleges and universities seek to demonstrate to current and future students (and their parents) the value of their tuition dollars. The institutional websites we reviewed illustrated a clear focus on the student perspective. This movement is no doubt a result of and a response to research that highlighted the top five educational practices and the importance of those practices to student learning, particularly as a means of attracting and retaining underserved student populations (see for example, Brownell and Swaner 2009). However, our study advances the view that institutional policy and practice lags behind institutional aspirations related to undergraduate research scholarship, and creative work. Implications for Practice The results of this study may be valuable to directors of undergraduate research programs and other institutional stakeholders. We offer three recommendations to improve practice. First, institutions should consider convening a task force to revise faculty evaluation procedures and promotion and tenure criteria to acknowledge faculty members’ mentoring of undergraduate researchers. Such a recommendation is not novel, but we believe it serves as an important first step in aligning promotion and tenure systems with institutional priorities and the changing, dynamic nature of scholarship to account for the variety of ways scholarship is enacted (O’Meara et al. 2015). We recognize the challenges of changing such procedures, but surely CUR member institutions should value this aspect of faculty work more highly than nonmember institutions. Despite the challenges, the importance of reconsidering promotion and tenure guidelines to acknowledge the work of mentoring undergraduate researchers cannot be underestimated. As Fairweather (2002) has argued, Bthe principal expression of academic values about faculty work lies in the promotion and tenure decision^ (pg. 27). Second, institutional leaders and directors of undergraduate research programs need to engage faculty members in strategic conversations about the types of support and recognition that mentors need when working with undergraduates. Very few of the websites analyzed linked faculty members to general resources and best practices in mentoring or strategies for mentoring students in under-represented groups, for example. This reality contributes to the ongoing conversation about the changing nature of faculty work (O’Meara et al. 2008; Simpson 1997) by shedding light on the disparities that exist as communicated through institutional websites, and in resources, incentives, and recognition available to students and faculty members. Our research suggests that institutional leaders may wish to consider routing resources directly to faculty members in the form of stipends, honoraria, and travel funds as incentives and acknowledgement of their efforts. Third, as we acknowledged in our Blimitations^ section, we only searched for explicit mention of undergraduate research, scholarship, and creative work search terms on the institutional websites. It is possible that faculty incentives and supports were more readily available at our sample institutions and that information was noted on other web pages or supported through other campus offices outside of undergraduate research. However, if that is the case, we argue that campus leaders and administrators need to ensure this information is more readily available to new and more senior faculty in the appropriate locations. Conclusion The Gallup-Purdue Index released in 2014 revealed that only 14 % of the 30,000 college graduates surveyed indicated they had contact with a professor who served as a mentor, caring about them as individuals, inspiring them to learn, and encouraging them to set lofty goals. Such data suggests that significant work lies ahead for leaders in higher education, especially as colleges and universities aspire to increase the availability and visibility of URSCWand the work that faculty do as mentors. Resources and policies, including promotion and tenure guidelines, need to be aligned so that faculty members are supported and recognized for their critical contribution to undergraduate research, scholarship, and creative work; and communication strategies in digital spaces should reinforce these efforts. Acknowledgments This project was developed through Elon University’s Center for Engaged Learning, which sponsored a multi-institutional research seminar on BExcellence in Mentoring Undergraduate Research.^ We also want to thank Christopher Morphew and three anonymous reviewers for their review and advice as we further developed this manuscript.

The establishment of an institutional repository has become a commonplace activity within academic libraries in the recent past, fuelled by the ready avail ability and relatively simple implementation of a number of open source software platforms and operating systems. However, the proliferation of repositories in academic libraries worldwide occasionally suggests that this trend is technology-driven rather than demand-driven. In general, the in-house development of a repository typically requires a time-consuming and labour intensive accumulation of bibliographic data and full text, and a concerted effort to either encourage or compel researchers to deposit pre-publication or otherwise copyright free copies of their work into the repository. To make the venture successful, this effort must be sustained long-term and becomes part of the mainstream activities of the Library. While the body of evidence supporting the value of open access repositories is growing, given the implementation costs and long lead times involved, university administrators might be forgiven, nonetheless, for questioning the eventual and direct benefit at the “micro” institutional level of investing in the development ofan institutional repository. While a repository might make published research available and accessible, does that necessarily translate into actual use and, therefore, make the institution's research more visible in a practical sense? The rhetoric of open access is that the presentation of the institution's published research output via a repository will inevitably improve its visibility and raise the profile of both the institution and the researchers involved, and potentially lead to greater citation frequency and impact of the published items. This appears to be supported by recent research on open access, such that placing research publications on open access leads not only to greater end user access but also to increased citation counts for the author, as well as greater potential for cross-disciplinary research. In the Hong Kong context, it has also been suggested that open access supports recent Government University Grants Committee (UGC) initiatives for the assessment of research impact, as well as the potential for know ledge transfer, through the greater accessibility of research output. In this paper, the development of an institutional repository by the Library of The Hong Kong Institute of Education (HKIEd) will be discussed in the context of pressures of time and production cost in the development of the repository and in terms of the immediately measurable impact that the repository has had. While many such studies might demonstrate the value of a repository over time, this study focuses in particular on the immediate effect after the repository was launched. Section snippets The Research Profile of HKIEd The Hong Kong Institute of Education was established as a specialist teacher education institution in 1994, through the amalgamation of five teacher training colleges and institutes. The Institute is celebrating its 15th anniversary this year and can also claim a seventy-year history of teacher education in Hong Kong, with Share Cite 1 2 one of its former colleges, the Northcote Training College, being established in 1939. The Institute currently has a student population of 6,385 students, across arange of Development of the HKIEd Research Repository In this context of a conscious Institute attempt to build research capacity and output, the Institute Library initiated the development of a “research repository” of staff publications. The Institute Library had previously developed an institutional repository which had concentrated on the digitization and organization of arange of “in-house” Institute publications, such as calendars, student handbooks, and commemorative publications. Hitherto, the function of the repository had been to serve Methodology for Repository Development Prior to the establishment of the HKIEd Research Repository, the public reporting of research publishing at the Institute had been largely through an author listing posted to the Research and Development section of the Institute Web site. This listing had a number of drawbacks which limited its effectiveness in promoting the Institute's research. Chief among these was that the listing could not be searched by any field, such as author, title or subject, therefore restricting its use to browsing Launch and Profile of the HKIEd Research Repository The soft launch of the Repository took place in late January 2009, when the initial database—consisting of approximately 9,000 staff publications—was first released for use by Institute staff. This comprised the entire research output of the Institute in the publication categories given above, from the establishment of HKIEd in 1994 through to the present day. Of these 9,000 citations, at the time of the soft launch of the Research Repository, it had been possible to locate and link to the full Use of the Repository to Promote Institute Research As mentioned earlier, in response to the Report of the UGC Review Group on the HKIEd Development Blueprint , a greater emphasis has been placed recently on improving the visibility of Institute research output. As the Research Repository is tailor-made for this purpose, Institute decided, shortly after the launch of the Repository, to create a link directly from the HKIEd Research Web site to the Research Repository, replacing the previous research listing of staff publications. This meant that Initial Response to the HKIEd Research Repository Within a month of the launch of the HKIEd Research Repository and its subsequent linking to the Institute's Research Web site, the Web traffic tool Google Analytics was applied to the Repository to provide a simple and no-cost means of tracking patterns of usage. In the subsequent two months of operation, this has produced some interesting and immediate indications of the level and type of usage. The Google Analytics data for April 2009 recorded that a total of 5,255 visits were made to the site Future Development of the Repository Since the launch of the Research Repository, a number of potential areas for improvement have been identified. The most significant of these has been a demand from academic staff using the Repository to archive and highlight their complete published output by including a range of “non-traditional” formats, such as musical scores, newspaper articles, speeches and so on, to reflect the entirety of the work that they do, as opposed to formal papers. In addition, with changes mooted to the faculty Conclusion The purpose of this paper has been to explore from the practitioner level some of the issues associated with the development of an institutional repository and to offer some insight into the impact that a repository may have on the visibility of published research. While the above analysis may indicate a modest initial take-up of the HKIEd Research Repository when compared to typical Library Web traffic overall, what it does suggest is that the making available of academic research through this.

A common interest in CSCL research is to study how different kinds of computer artifacts can and do scaffold learning as part of collaborative activities. This mutual reference point inevitably directs our analytical attention to discourse, simply because it is the most important medium through which thinking develops and is made observable (Mercer, 2000). Here we use discourse as a generic term indicative of all forms of talk and text. For this reason, it has been an important aim for many CSCL researchers to design for and investigate the forms of discourse that are crucial for the development of thinking (see, for example, Hakkarainen, Lipponen, & Ja¨ rvela, 2001; Mercer & Wegerif, 1999). This research has generated many important insights into the structures and functions of discourse that are beneficial for learning, including how CSCL artifacts contribute to structuring such discursive practices (Baker, Hansen, Joiner, & Traum, 1999; Edelson, Gording, & Pea, 1999; Muukonen, Lakkala, & Hakkarainen, 2005; Scardamalia, Bereiter, & Lamon, 1994; Roschelle & Teasley, 1995; Suthers & Hundhausen, 2001). Seeing that, how the institutional contexts into which CSCL tools have been introduced actually impact their use has not been a particularly important topic in CSCL research despite the recognized centrality of discourse (see Lipponen, 2001; Arnseth, 2004). This provides us with a rationale for critically examining this issue in more detail. In order to grasp the institutional contexts of CSCL activities, we need a certain conception of how the relationships between discourse, learning, and technological tools on the one hand, and the context in which they are used on the other, can be conceived theoretically and pursued analytically. Our argument concerns how this relationship has been and perhaps should be conceived in CSCL research. The point is that differences in analytical practices have consequences for the generation and assessment of findings, for what we consider to be productive in terms of learning, and for how we as researchers can contribute to fostering the development of educational practices. However, it is also important to consider whether there might be any points of convergence across approaches, particularly relating to normative criteria for fostering effective CSCL environments. In order to structure our argument, we introduce a distinction between what we term systemic and dialogic approaches to CSCL research (Dillenbourg, 1999; Linell, 1998). This distinction cuts across any neat separations between theories commonly employed in CSCL research—such as sociocultural or cognitive theories of learning and thinking—in that it directs our attention more explicitly to issues dealing with methodology and analytical practice. For example, even though many studies claim to adhere to more social and cultural approaches in theory, how the meanings and functions of CSCL tools are actually constituted in practice are rarely demonstrated analytically (Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003). As we will show, employing this distinction enables us to make observable how particular aspects of our object of inquiry—participants interacting with or through tools in an organised setting—is either made available or unavailable for analysis. In order to provide a general context for our argument, we will first highlight some key findings in the CSCL field without necessarily discriminating between the technological tools, theoretical perspectives, or methodological designs employed. In the second section, the distinction between systemic versus dialogic approaches is defined and worked out in more detail. In the following two sections, we have chosen a few significant studies in order to work through the analytical and methodological argument being put forward. We then analyze an excerpt of data from our own research in order to provide a practical demonstration of the usefulness of our approach. Finally, we discuss the implications of the different approaches including how together they might contribute to establishing a more well-founded body of knowledge as regards the effectiveness of CSCL. A brief overview: Successes and failures of CSCL Numerous CSCL studies demonstrate the positive effects of CSCL tools on the quality and amount of social interaction and other features of the teaching–learning process. Moreover, theoretically derived hypotheses regarding the impact of information and communication technology (ICT) supported collaboration on learning have been supported by empirical evidence (Lehtinen, Hakkarinen, Lipponen, Rahikainen, & Muukkonen, 1999). For example, CSCL is reported to facilitate task orientation and reflective activity (Cohen & Scardamalia, 1998), reasoning and argumentation (Hoadley & Linn, 2000), mathematical problem solving (CTGV, 1997), student’s beliefs about the nature of learning (Hewitt, 2001), and the learning of complex scientific concepts and processes (Roschelle, 1992). Also, in comparative studies of CSCL and non-CSCL students, it is reported that CSCL students outperform non-CSCL students on standardized achievement test scores in mathematics and reading (Lamon, Secules, Petrosino, Hackett, Bransford, & Goldman, 1996). Furthermore, CSCL is reported to support collaborative knowledge building, including progress in developing deeper understanding, generation of further questions for inquiry, and engaging in collaborative discourse to advance explanations and arguments (Edelson, Gording & Pea, 1999; Scardamalia, Bereiter & Lamon, 1994). In the same vein, it can support students in their establishment of shared understanding as part of collaborative problem solving (Baker et al., 1999; Roschelle & Teasley, 1995; Suthers & Hundhausen, 2001). CSCL is also said to facilitate student’s meta-cognitive understanding (Brown, Ellery, & Campione, 1998). To summarize, these studies demonstrate quite clearly that different types of CSCL tools under certain conditions can be a part of practices that produce more effective and productive learning outcomes. However, disadvantages with CSCL are also reported. Regardless of whether CSCL is used in distributed or co-located environments, lack of discussion, argumentation, and challenging of ideas are common findings (Guzdial, 1997; Hewitt & Teplovs, 1999; Lipponen et al., 2003). This is particularly the case when CSCL tools have been introduced into ordinary classroom settings. In these cases, activities have generally been centered on knowledge reproduction and on producing acceptable outcomes with the least collaborative effort. Moreover, ambiguity, disagreements, or diverging ideas are seldom resolved in any productive manner (Arnseth, 2004; Lipponen, 2001). Consequently, it is problematic to make the positive results reported above more generally relevant across contexts. According to Lipponen (2001): Although the new technology and the theoretical and pedagogical ideas support each other, the attempt to promote educational use of CSCL technology, and at the same time implement new pedagogical and cognitive practices of learning and instruction, appears to demand the utmost of both teachers and students. Many of the technical, theoretical, and pedagogical insights have not been transformed into widely adopted practices of teachers and students (p. 11). However, as Lipponen (2001) is careful to point out, these rather disappointing findings (at least from a normative point of view) cannot necessarily be attributed to the nature of CSCL tools as such. On the contrary, the failures of technological tools to produce the proposed effects, including the pedagogical models underpinning their design, need to be examined in relation to the context in which they are used (Arnseth, 2004; Ludvigsen, in press). However, as will become clear, context is by no means an uncontroversial concept. Still, in regard to these failures, it also seems necessary to remind ourselves that the majority of CSCL studies conducted in ordinary classroom settings are design and intervention studies, meaning that they are usually carried out over a few days or perhaps weeks at the most (see also Hakkarainen, Lipponen & Ja¨ rvela¨ , 2001). Consequently, the CSCL tools in question have not become an integrated part of the long-term development of institutional practices (Wasson, Hoppe, & Ludvigsen, 2003). As a result, the existing features of schools—teaching practices, evaluation practices, or technological infrastructures—are seldom taken into consideration in accounts of findings. If they are referred to at all, they are generally conceived as internalized norms serving as explanations of failures, e.g., that the teachers and students had different goals than was implied by the CSCL tool in question (see Hewitt, 2001). As we will demonstrate in more detail below, however, the actual enactment of these practices has important implications in regard to the effects that CSCL tools might have. To summarize, disagreements between the approaches suggested above mainly concern how the institutional context should be understood and identified within the analytical schemes employed. However, before we go any further, it is necessary to provide more detailed definitions of the approaches we are proposing. Systemic versus dialogic approaches to CSCL In order to simplify, we might say that a fundamental tenet of research adhering to a systemic approach is its attempt to generate models of how specific features of technological systems affect collaboration, reasoning, functions, contents, and structures of discourse (see, for example, Dillenbourg, 1999). The analytical purpose is not necessarily to develop causal models, but rather to identify the interdependencies between different variables, including how specific features of the technology facilitate students’ understanding or ability to solve problems in a variety of knowledge domains (Salomon, 1993; Kirschner, Martens, & Strijbos, 2004). The task for the analyst is to describe and account for the configurations of elements that are most beneficial in terms of some outcome measure of what has been learned. That is to say that the analytical focus is on describing the systematic relations between forms of social interaction, and specific types of support or other contextual factors on the one hand, and qualities of outcome on the other. The result of such an analytical practice is the formulation of a model, or the readjustment of a previous model, which specifies the correlations between the variables that were defined at the outset and inscribed into the analytical scheme employed. Such a model might state that a CSCL application, together with certain language practices, e.g., requests for clarifications, together are likely to produce positive learning outcomes (see, for example, Roschelle & Teasley, 1995). In accordance with this approach, the institutional context would mark something that surrounds the activities in question and that constrains or perhaps facilitates in specific ways what the participants do (see also Cole, 1996). Having said that, institutional norms and rules are also internalized by teachers and students, and they can, for example, be identified through the use of questionnaires or interviews. Furthermore, their (cor)relations with specific technological affordances or outcome measures can be determined through statistical analysis. In research adhering to a dialogic approach, on the other hand, the focus is on how the meanings and functions of discourse, tools, and knowledge are constituted in social practices (Sa¨ ljo¨ , 2000). According to Linell (1998): ...dialogism regards every cognitive and/or communicative act as an Banswer,^ as responsive to something (often only implicit) in the contexts. A contribution to dialogue, whether a single utterance or a lengthy spate of talk, is made coherent by being related to some (often implicit) issue (Bquaestio^) of current relevance; the contribution must be rendered accountable (by the actor or the analyst) in relation to the ubiquitous metaquestion Bwhy that now (to me etc.)^ (pp. 35–36). The meanings and functions of one variable cannot be treated as distinct and separable from the others. On the contrary, the different elements mutually shape one another, and their meanings and functions are results of local negotiation and sense making. Thus, rather than being separable nodes in a network of relations, they become mutually laminated onto one another in and through social interaction. As such, social interaction with artifacts in an organized setting becomes the site where these processes are made available for study (see also Middleton & Brown, 2005). Thus, in order to understand how CSCL tools, pedagogical models, and knowledge are made sense of, including their possible effects on the pedagogical practices in question, we as analysts need to carefully scrutinize the sequential unfolding of activities along different time scales (Lemke, 2000). This is because any action is responsive to what happened before and at the same time it projects possible responses in the future (Linell, 1998). Therefore, instead of treating social interaction as a relatively neutral intermediary between cognitive and external contextual variables, it is brought into the center of analytical attention (Sa¨ ljo¨ , 2000; Wells, 1999; Wertsch, 1991). It is here that the meanings and effects of CSCL tools become available for study. Of course, this kind of research can also identify genres and structures whose general relevance goes beyond the immediate situation, as well as being able to construct models of the kind of CSCL uses that are likely to be most effective and productive. However, instead of treating models as explanations of and templates for action, they are conceived as resources for action (Suchman, 1987). That is to say, their potential usefulness is established in dialogue with other features of the setting that the participants need to manage as part of their day to day activities (see Kvale, 1996; Rystedt, 2002). In terms of how the institutional context is understood, the principal analytical ethos is to start with examining what students and teachers actually do (Sa¨ ljo¨ , 2000). This does not rule out any concern about examining the historical genesis of the artifacts or practices in question or the specific institutional arrangements having to do with technological infrastructures, division of labour, or specific institutional rules and regulations (see Ma¨ kitalo & Sa¨ ljo¨ , 2002). The point is that this contextual framework is not seen as determining local practices. On the contrary, they are actively oriented to, reproduced, or resisted in and through action (Arnseth, 2004). Still, depending on the unit of analysis and level of description preferred, either individual’s changing participation in dialogue or institutional orchestrations of learning could be highlighted in the actual analysis (Valsiner, 1994; Valsiner & Van Der Veer, 2000; Ludvigsen, in press). To summarize, the aim is not to understand how different variables covariate, but rather to understand how the meaning of knowing, knowledge and artifacts is constituted in dialogue between participants, who through their actions are responding to various contextual features of the setting and are thereby making them relevant. After having provided more elaborate definitions of the approaches, there is a need to demonstrate their consequences for analytical practice more clearly. A systemic approach to CSCL research In order to provide a detailed critique of research belonging to the systemic approach, we will focus primarily on discussion and inquiry types of CSCL applications. The reason for limiting ourselves to these kinds of tools is partly practical. Nevertheless, in formal learning institutions, applications of this kind have existed for some time and they are generally available for use outside of design projects that are rather limited in terms of scope and dissemination. Computer-Supported Intentional Learning Environment (CSILE), for instance, was one of the first applications designed to support collaborative learning. Moreover, together with its various implementations such as Knowledge Forum and WebCSILE, it is one of the few applications that has been widely used and tested in ordinary educational settings over longer stretches of time (Miyake & Koschmann, 2001). Therefore, to provide a critical discussion of some of this research seems particularly relevant because the tool might have become more attuned to developing institutional practices. In addition, according to Lehtinen et al. (1999) there is substantial empirical evidence for the fact that CSILE facilitates higher-order cognitive processes, regarding, for example, the ability to read difficult texts, the quality of developed questions, and the depth of explanation and problem solving in mathematics. Still, there is a need to unpack this evidence in a bit more detail. For example, Hewitt (2001) did a comparative case study of two grade six Human Biology units, each taking place over 6 weeks, where, each day, thirty minutes were allocated for work with CSILE and thirty minutes for research. In his analysis he relies on an interview with the teacher in addition to the content of the CSILE database. The first unit represented the teacher’s initial efforts to develop a knowledge-building community, while the second took place two years later. According to Hewitt (2001), the teacher had by this time developed instructional strategies that were closer to the normative pedagogical ideal embedded in CSILE. The results from the first unit were disappointing (Hewitt, 2001). First, even though students followed the teacher’s instructions there was a lack of collaboration. Second, there was a lack of conjectures, meaning that the students rarely shared their theories and assertions with others. Third, the plans that the students produced were weak and focused on topics rather than process. Fourth, the gathering of information was poor, meaning that students examined broad areas rather than specific problems, which resulted in a gradual accumulation of knowledge without any discrimination. Fifth, the students produced too many questions that were left unanswered and, moreover, they rarely referred back to their questions during their activity. Therefore, the questions played a minor role in structuring the activity. These findings resonate with the ones reported above relating to CSCL applications that were introduced into ordinary educational settings. To assess collaboration, Hewitt examined each note in the database in order to determine whether it explicitly or implicitly referred to other notes. Only 15% of the notes fitted this rating, and about two thirds of this particular collection of notes was considered superficial in content. However, Hewitt does not provide any criteria for categorizing a note as being collaborative or not. Thus, it is not made explicit what is entailed by the categories implicit and explicit. According to Hewitt (2001) this lack of collaboration might reflect that the students not understanding the nature and purpose of CSILE (Hewitt, 2001, p. 23). According to him, the students B...seemed to perceive the program as an environment for project-based work where their main objective was to seek out and replicate information from texts^ (2001, p. 23). Even though the explanations provided by Hewitt seem very reasonable, he provides no evidence concerning how the activity proceeded. As such, the inferences about student perception of CSILE is simply asserted rather than demonstrated analytically. As we will show, this is an effect of the analytical scheme employed. In contrast to the first unit, the second Human Biology unit fit the goals of CSILE to a larger extent because by this time the teacher had developed a set of strategies for facilitating discussion (Hewitt, 2001). The analysis of the database showed that the number of collaborative entries increased from 15% to 43%. Moreover, the percentage of messages rated as conjectures—messages that contained the tag My Theory—rose from 1% to 37%. Hewitt concludes that the change in activity patterns was mainly due to the fact that the teacher changed the focus from task completion to developing understanding. This is a very interesting finding indicating that when CSILE becomes more attuned with developing teaching practices, it is used more productively and effectively. However, the change of instructional practices is inferred only on the basis of an interview with the teacher. What is more, a change in student reasoning and problem solving is inferred on the basis of a specific thinking-type tag attached to their messages. To summarize, Hewitt identifies a change in activity patterns, a change that he attributes to the development of teaching practices. This change in activity patterns is again linked to more productive reasoning. As such, even though he does not provide any correlational analysis, his research strategy is to describe a set of systemic relations. The development of teaching practices is a result of the fact that the teacher is able to align his practices with the CSILE design, and this is treated as an effect of his ability to internalize the CSILE pedagogy. Thus, how the changes in practice develop in tension or in conjunction with the institutionally appropriate and authorized ways of doing learning and teaching is treated as analytically uninteresting by fiat. In our view, these are general problems with studies that use content analysis of a CSCL database combined with interviews, surveys, or social network analysis as the only sources for making inferences about changes in teaching and learning practices (cf. Lipponen et al., 2003). That is to say, the nature of teaching and learning is predefined at the outset and, by the same token, how participants themselves actively establish contexts for learning is simply disregarded as analytically uninteresting (Jordan & Henderson, 1995). Thus, even though a systemic research strategy makes it easier to determine correlations between variables and to make systematic comparisons across datasets, it makes us miss on crucial aspects of the key object of inquiry for CSCL research. A dialogic approach to research on CSCL In order to demonstrate a dialogic approach, we will briefly address a few of the most relevant studies. The studies are also selected in order to illustrate differences in analytical practices within a broader dialogic framework. According to Stahl (2001), studies embedded in such a framework have not been particularly prominent in CSCL research (but see more recently Arnseth, 2004; Ivarsson, 2004; Ludvigsen, in press; Stahl, 2006). The types of ICT applications used in the studies discussed below are not necessarily comparable with one another or with discussion and inquiry types of tools. Neither have their use within the institutions in question been cultivated over long stretches of time. However, in this context we believe this is not a major problem as our aim is to compare systemic and dialogic research practices and not empirical findings as such. Still, for future research it is crucial to pursue dialogical research strategies over longer stretches of time in order to determine how productive uses of CSCL tools actually develop. At this point, such research designs are very rare. In accordance with a dialogic approach, CSCL applications are not treated as a variable where their relationship to other variables can be determined statistically. On the contrary, the analytical concern is with how computer applications provide a context for social interaction. Important contributions in this regard have been made by Mercer and colleagues (see, for example, Mercer, Phillips, & Somekh, 1991; Mercer & Wegerif, 1999; Mercer et al., 2003). For them, thinking is conceived as a form of communication where knowledge forms part of what the talk is about; that is, it becomes part of arguments, disputes, explanations, clarifications and so forth (Mercer, 2000). In their research, they have put considerable emphasis on making what they term the ground rules for talk explicit to learners. Put simply, ground rules refer to Bthe implicit norms which govern the spoken interactions between teachers and pupils, and which generate its familiar and distinctive patterns^ (Mercer et al., 2004, p. 4). According to them, exploratory talk is particularly productive for the development of joint thinking (Mercer & Wegerif, 1999). Exploratory talk is characterized by the mutual development, discussion, and reflection upon ideas and problems. Furthermore, it is a continuous and mutual accomplishment by participants engaged in collaborative activities. In a comparative case study of collaborative activities involving the use of an educational computer program called Kate’s Choice, Mercer et al. (2004) found that the discourse of the class where exploratory talk had been nurtured as part of a specific program was very different from that of a control class where the same software was used. Kate’s Choice is a kind of interactive narrative, designed in order to facilitate moral reasoning. According to Mercer et al. (2004), the children asked one another task-focused questions, provided reasons for statements and challenges, considered several positions before making decisions, and agreed on a solution before acting on the computer program. In contrast, in the control groups the child controlling the mouse made decisions without consulting others in the group, the choice of the most dominant child was usually accepted, arbitrary decisions were made without considering alternatives, and children spent very little time on each decision before moving on to the next step in the program (Mercer et al., 2004). Interestingly, the findings generated in the control group are similar to the ones mentioned above concerning the introduction of CSCL applications into ordinary classrooms. Their findings suggest that if computer-supported collaborative group work is complemented with certain language practices, the computer provides a good framework for collaborative learning (Mercer et al., 2004). That is to say, students used the prompts made available by the tool as an opportunity to engage in exploratory talk. However, as Mercer and Wegerif (1999) are careful to point out, the fact that the tool supports learning is not due to its design as such, but to the language practices in which it is entrenched. Mercer and Wegerif (1999) argue that exploratory talk is an analytical category that they find useful for examining the relationship between talk and thinking. However, as they themselves acknowledge, it is not always easy to distinguish between different forms of talk in practice. Therefore, we as analysts still face the practical problem of identifying exploratory talk in what students and teachers do. By employing such a category we might miss out on how the talk actually emerges and how different aspects of language use co-constitute exploratory talk. Moreover, it makes it difficult to examine the diverse ways that participants’ actions are produced in response to certain normative orderings made relevant by the situations in which they act. Therefore, focusing exclusively on productive talk and interaction makes it difficult to analyze how developing discursive practices also demand changing institutional practices. In contrast, Crook and Light (2002) make institutional practices into a focal point for study in regard to the challenges involved in facilitating learning with ICT. Their concern is with the dynamics between everyday practices and the practices of study, something which is made relevant when students enter into their first year at university. An important question concerns whether computers might serve to remediate more traditional modes of academic communication, such as lectures, seminars, and tutorials. According to Crook and Light (2002), in order to facilitate processes of enculturation into academic practices, universities provide scaffolds that sustain activities such as: ...engaging with exposition, orchestrated discussion, research, systematic annotation, the focused reading of text, and a variety of other directed activities that many students may not find easy to mobilize and manage independently (p. 174). According to them, these practices of formal study are closely interrelated to practices that students are familiar with and which are well rehearsed as part of their everyday life. In short, they find that developing new practices with ICT is very difficult, something which is not due to students’ lack of familiarity with the technology. On the contrary, they argue that the tools and their associated practices are not particularly well attuned to already existing practices. In regard to virtual seminars, for example, they report that the productivity of the interaction was dependent on whether the discussion was extensively moderated by tutors (Crook & Light, 2002). On the other hand, the asynchronous character of the interaction did not seem very productive for students. The authors argue that this was because it is too different from talk in seminars which, according to them, often go well because they are grounded in the everyday practice of speaking. However, they are careful to point out that the medium is not intrinsically problematic and such practices might become productive over time. However, this is dependent upon them being cultivated as part of various institutional practices, e.g., doing web-based tutoring on students assignments, etc. Their basic argument is that formal learning can be very difficult, but that this is made easier by the fact that formal learning emerges out of practices with which students are already familiar (Crook & Light, 2002). The relevance of this study in regard to our argument is that their analysis makes visible how the productivity of computer-supported activities is dependent on its fit with already established institutional practices. Thus, developing productive CSCL environments also entails changing institutional practices—the institutionally appropriate ways of doing teaching and learning. Still, a problem with Crook and Light’s (2002) study is that they do not provide any detailed analysis of how these practices actually converge, diverge, or are enacted in social interaction. In order to develop our argument even further it is therefore necessary to provide an analytical example in order to demonstrate how the relations between knowledge construction, computer artifacts, and institutional concerns can be analyzed in some detail. In our example, the institutional context is analytically accounted for through references to what the participants display an orientation to and manage in and through their actions. An analytical example from our own research The excerpt we analyze is taken from the DoCTA NSS project (Design and Use of Collaborative Telelearning Artifacts, Natural Science Studios). In this project we introduced the CSCL application Future Learning Environments 2 (FLE2) into a classroom setting and we adopted the progressive inquiry model (PI) as the main design principle (Muukonen, Hakkarainen, & Leinonen, 2000). The categories the students were supposed to use in their knowledge construction were: problem, my working theory, reliable knowledge, uncertain knowledge, comment, meta-comment, and summary. These categories are modified versions of the ones found in FLE2, which is a discussion and inquiry type of CSCL application similar to CSILE (Ludvigsen & March, 2005). Generally, students displayed a certain difficulty with categorizing their notes both in terms of what categories they should use and how the categories could be used as scaffolds for the development of their arguing and understanding (see Arnseth, 2004). In excerpt 1, the three girls—Sara, Anne, and Lene (S, A, and L in the excerpt)— who are all sharing one computer, are talking about what kind of category they should use as a description of a message in FLE2. The episode happened at a stage in their activity when they were engaged in knowledge building and used the categories embedded in FLE2 in order to develop their arguments. How might we go about identifying parts of the context toward which the girls are displaying an orientation to in this excerpt? For example, in what ways is the CSCL application part of this context, and what is more, how can we identify particular institutional responsibilities having to do with how they deal with knowledge? What are the challenges in terms of developing more productive practices? Of course, in order to make substantial claims, we would need to examine how students used categories across groups and over time. Moreover, in order to make sense of their talk, we would also need to know something about the tool and the pedagogical ideas embedded within it. In this instance, this excerpt is used as a resource for illustrating a particular analytical practice. In the first few lines of the excerpt Anne and Lene disagreed on how they should categorize a particular knowledge object. In line 1, Anne expressed that they should use the category reliable knowledge while in line 2 Lene responded by saying that they should employ the category insecure knowledge. Anne produced another disagreement token in line 4. The particular knowledge object they were discussing is an excerpt of a newspaper interview with a professor who is sceptical of the use of gene testing. In line 5 Sara joined their discussion and displayed agreement with Anne. We can infer that the CSCL application structures their interaction in at least two senses. First, the practical task that the students encounter, which is to categorize a fragment of knowledge they have found in an additional text, is made relevant by the categories in FLE2. Thus, the system of categories that is inscribed in the technological system makes certain actions relevant on the part of the users (Goodwin, 1997). However, it does not necessarily imply that they offer reasons for their choice, or to put it differently, the application does not determine how they go about categorizing. As such, choosing a category is very much a practical problem for the students. Furthermore, whether they are able to use the categories as scaffolds in their activity is, among other things, dependent on whether they challenge each other’s ideas and whether these challenges are taken up and responded to by others. Second, the tool makes available a whole set of categories, and it is not easy to distinguish between them because they do not mutually exclude one another. This is because any knowledge object can be categorized in a number of different ways. In line 8, Lene provided a reason for her claim stating that they should not use the category reliable knowledge Bit is not eh just because he says so.^ This account is interesting because it questions whether the validity of a statement should be assessed simply on the basis of the authority of the person who claims it. As such, it is an account that questions uncritical copying of knowledge from authoritative sources. In dialogical terms, this account could be a starting point for a more elaborate discussion of the epistemological status of the texts that the students were going to use in order to substantiate their claims. However, the opportunity for elaboration that is made available by this account was not taken up in the following talk. Sara did not challenge Lene to explain why it should not necessarily be treated as valid knowledge in line 9. Instead she suggested a different category. Another category that is available and which also might be considered relevant is uncertain knowledge. Here Sara simply readjusted her position in their joint discussion and displayed agreement with Lene’s previous accounts in lines 2 and 6. That is to say, Sara inferred that the text in question should be given a label which is consistent with Lene’s critique. The category insecure knowledge is an available category that can accommodate this critique and still be able to do the work required. Lene provided a more elaborate reason in line 10 where she stated that it is B(...) different when it says it was eh a survey.^ An important point that needs to be emphasised is that deciding upon a category is connected with the practical management of disagreement within the group. Thus, the choice of the category insecure knowledge might enable the group to manage disagreement, something which is an important concern for participants in collaborative encounters such as this. However, even though this category might enable them to solve this particular problem, it is not treated as adequate by Anne who offers a set of alternatives in lines 11–13. At first she suggested Bthe blue one,^ which is uncertain knowledge. However, she also offered two other alternatives; Bthe white one,^ which is meta-comment, and the category comment or Bprocess commentary^ which is the exact formulation she employed. She provided a reason for offering these alternatives in line 15. Even though the system of categories structures their interaction, the students make use of them to manage their practical concerns, which in this case was to categorize a piece of knowledge, taken literally, without necessarily providing any reasons for why they have selected a particular category. We might say that they understand their task as involving the collection of arguments and to categorizing them in accordance with the template of categories made available by the artifact. Moreover, the tool is interpreted and constituted in order to fit this concern, which is about how they can go about finding facts that support their case and whether these facts actually qualify as facts and can be given the category reliable knowledge. However, in this particular case, this concern conflicted with internal group dynamics. In general, there was a preference for this category, because it does some important work. On the one hand, it qualifies their accounts as credible and as more robust against rebuttal (see Arnseth, 2004). On the other hand, it also labels their accounts as being in accordance with their task, which was to produce reliable accounts grounded in authoritative knowledge. By employing this category students are able to manage their responsibilities for doing institutionally relevant actions. However, in order to analyze how students made sense of their task, we would also need to look into how the task is introduced and, subsequently, how this is taken up or resisted by the students. As such, the actual meaning of the task would be an effect of local negotiation. Institutionally relevant actions are not fixed and immutable. On the contrary, even though the teacher did not challenge them here, there might be other episodes where he or she could request an explanation of why their arguments were reliable. Still, there is no guarantee that the students would take up and respond to this challenge. They might use evasive strategies and argue that they had done their task appropriately (Arnseth, 2004). In this excerpt, through some form of minimal collaborative effort, a choice of category was made in line 18. The category they ended up with was the category commentary, a category which was not disputed by anyone within the group. However, it is also a category that was not necessarily relevant. This is due to the fact that the category commentary should ideally be a comment regarding the development of their knowledge-building activity. However, in this case the preference for agreement within the group made them use this category since it was uncontroversial. Against this background, we can see that developing teaching and learning practices with CSCL tools is by no means straightforward, as a number of interrelated factors constitute such practices. That is to say, it is dependent on whether the students are able to make sense of the tool and see it as relevant. Moreover, the teacher needs to challenge the students and help them to make sense of their task, including how the tool might facilitate their work. As we have shown, the meanings and functions of the application are by no means self-evident to students. This is closely intertwined with what is constituted as institutionally appropriate ways of dealing with knowledge. The concern that the participants in this excerpt were dealing with was to decide and agree upon a category that, for all practical purposes, could be used as a description of their note. This task was institutionally embedded, in the sense that they were accountable for doing the task in a particular way. Institutionally there was a preference for the category reliable knowledge, that is to say they were supposed to develop their arguments so that they became more valid. However, as shown in the excerpt, this task was intertwined with internal group dynamics and issues having to do with the management of disagreement. The initial disagreement within the group was dissolved by invoking an uncontroversial category that the group could agree upon, but which, from a normative perspective of knowledge building, was not necessarily a relevant description of their note. However, in this case the preference for reaching agreement took precedence over the need for understanding the relation between some category and a knowledge object. Concluding remarks and future steps At this point in the development of CSCL as a field of research it is reasonable to ask whether there are any possibilities for convergence across approaches. We do not believe that the approaches discussed here can be reconciled in any simple sense. On the contrary, as we have demonstrated, they are to a certain extent incommensurable as they pursue very different analytical strategies. However, this does not mean that they cannot learn from one another. Acknowledging their differences, the approaches might inform one another in providing directions for future CSCL research. That is to say, research belonging to the systemic approach provides important findings in terms of what works and what does not across contexts, including how the effectiveness of CSCL applications might be systematically related to the previous experiences of students and teachers or features of the institutional context. However, this kind of research does not provide any detailed information on how changes in teaching and learning practices actually come about and are negotiated in dialogue among participants responding to various normative features of the setting. In order to analytically make sense of this, there is a need to examine the sequential unfolding of activities along different time scales. Consequently, in order to gain further insights into the complexities of CSCL, we need both approaches, but it is important to keep in mind that they are useful for different purposes since they make different aspects of students’ CSCL activities available for study. Having said that, there is a need to spell out in detail what the points of convergence between the approaches might be. In and through our discussion, we believe that we are able to identify two fruitful points of convergence, discussion, and argument across these approaches. As mentioned previously, in their comprehensive review, Lehtinen et al. (1999), argue that there is substantial evidence for the fact that CSCL environments, under certain conditions, bring about knowledge-seeking patterns and higher-orderthinking skills. However, these findings have not been replicated when CSCL tools have been introduced into more ordinary classroom settings (Lipponen, 2001). Consequently, Lipponen (2001), for example, argues that the productivity and effectiveness of CSCL applications is closely related to social and cultural aspects of the settings in which they are introduced. In the same vein, but focusing more explicitly on processes of cultural transformation, Crook and Light (2002) demonstrated the complexity involved in developing the cultural practices of learning and teaching with the support of ICT at the university level. As such, research demonstrates quite clearly that there is a need to take the institutional context into account. This constitutes one point of convergence. On the other hand, Crook (1998), for example, emphasized that CSCL environments can be beneficial for learning if students articulate their thinking and express their ideas so that limitations in understanding become accessible and publicly available (see also, Arnseth, 2004; Krange, in press; Rasmussen, 2005). This is a necessary process for revision and elaboration of ideas (Ludvigsen & Mørch, 2003, 2005; Stenning et al., 2002). Thus, regardless of the particular approaches employed, there seems to be some shared understanding of what it is that affords learning, regardless of whether learning is conceived as Binternalization^ or Bchanged participation in social practices.^ In the learning sciences, many different concepts are used to characterize such productive learning practices: for example overcoming cognitive conflicts and epistemological break downs, (re)framing of the activity at hand, re-establishing mutual understandings, responding to challenges by teachers and fellow students, and joint exploration of problems. All of them point to the fact that students need to engage in transformative dialogue in order to develop more advanced problem-solving, reasoning, and arguing in regard to relatively complex curricular content. Put simply, disagreements concern why it is that certain forms of collaboration and discourse are considered to be beneficial for learning. Despite these disagreements this constitutes the second point of convergence. As we have demonstrated, however, these two points are closely related. That is to say, the historically developed practices of education are constitutive for the meaning and function of CSCL tools. Scardamalia and Bereiter (1996, p. 252; see also Hewitt, 2001) identify four characteristics of schooling that inhibit the development of student expertise. First, schooling still remains focused on individual student learning. Second, schooling deals mainly with demonstrable skills and formal knowledge that students are expected to memorize. Moreover, it is mainly the teacher who organizes lessons, asks questions, and summarizes activities. Third, to a great extent the learning objectives remain invisible to the students. That is, they are transformed into specific tasks and the procedures for accomplishing those same tasks. Fourth, the organization of the exercise of expertise is available only to the teacher and no mechanisms are provided for passing on the teacher’s expertise to students. That is to say, educational practices are still grounded in a transmission model of learning and on a mind-as-container metaphor (Lakoff & Johnson, 1980), where, more or less, it is the individual student’s responsibility to make sense of the teacher’s instructions (Sa¨ ljo¨ & Bergqvist, 1997; Sa¨ ljo¨ , 2000). As such, facilitating learning with CSCL also entails changing these institutional practices. In order to understand how such changes come about, we also need to pay attention to the sequential unfolding of activities in time. Our main argument is, therefore, that we need to examine more closely how the meaning and functions of CSCL applications are actually constituted in practice. In the CSCL community, research adhering to a dialogical framework can provide fruitful accounts for the temporal dimensions of learning and knowledge construction. As we have shown, this is crucial for understanding why CSCL applications fail or succeed. Paying close attention to the sequential organization of interactions might also enable us to understand how we can better facilitate learning with CSCL, in the sense that we can generate systematic knowledge about the forms of support that are likely to have the proposed effects on student talk and actions. However, in line with the dialogical approach, these effects are not infallible. On the contrary, they need to be (re)produced in and through social interaction. Transcript symbols = absence of a discernible gap (.) short pause ... untimed pause (...) omitted or inaudible talk ? marks rising intonation , continuing intonation [ ] clarifying information Acknowledgements An earlier version of this manuscript was presented at the first Kaleidoscope CSCL SIG symposium in Lausanne Switzerland, 2004; we would like to thank the participants for valuable comments. We also would like to thank the ijCSCL reviewers and the editors for valuable comments and critique. This research is supported by the Norwegian Research Council through grants to the following projects: BLiteracy and media convergence. New forms of semiotically mediated interaction^ (Grant nr. 157485/510) and TRANSFORM (Grant nr. 161946/S20). It is also supported by the strategic research effort BCompetence and Media Convergence (CMC)^ at the University of Oslo (http://cmc.uio.no/).

The purpose of this study was to identify the major science education programs in the United States where scholarly activity is published in science education research journals. This is the first study of the research productivity of science education programs at institutions of higher education, thereby, addressing Brown’s (1969) assertion that periodically scholarly fields should ‘‘take a look at itself’ (p. 263). Publications in science education research journals reflect the interests of the authors, the attainment of professional standards for scholarship that includes reviewers and editorial board, and core values of science education as a discipline. In Australia, Print and Hattie (1997) reported that refereed research publications are the best indicator for determining institutional/program quality. A second purpose of the study was to contrast two methods of institutional productivity raw and weighted count. Thereby, the study is providing a foundation for subsequent research which may include charting the shifting terrain of science education research. The science education research journals chosen for this analysis were: Journal of Research in Science Teaching, Science Education, International Journal of Science Education (only U.S. authors), Journal of Science Teacher Education, School Science and Mathematics (only science articles), Journal of Computers in Mathematics and Science Teaching (only science articles), Journal of Science Education and Technology (since 1992, when it was first published), and Journal of Elementary Science Education. We concentrated upon these journals which publish empirical studies emphasizing science education. It is acknowledged that some science educations journals, (e.g. National Science Teachers Association journals, Physics Teacher, American Biology Teacher, etc.) sometimes contain research articles, but their main emphasis is on teaching strategies and/or curriculum development (White 2001). It is acknowledged that additional science education research was published in other journals. Therefore, the total research productivity of science education programs could be considerably greater. According to Smith et al. (2003), research ‘‘productivity studies are controversial because questions about program quality are implicit in the ranking of institutions... criticized as little more than academic horse races... [however], because they are indicative of the extent to which programs... are contributing to the advancement of knowledge within a given discipline’’(p. 423). They concluded productivity should focus on research at the institutional level in identified research journals of the field. Related Literature Perception Research Institutional productivity studies are more common than individual scholars. Institutional program areas have been ranked based upon reputation and other factors than research publications. For example, U.S. News and World Report has an annual ranking of best graduate programs, including education. However, these rankings do not break education into specific areas, such as science education. An early study by Carter (1966) had ranked faculty by their recognition within particular departments. Goldberger et al. (1995) modified this approach by providing select faculty a list of publications for each program being evaluated. Lowry and Silver (1996) based their ranking of political science departments on the reputation of both the department and the institution. West and Rhea (1995) compared programs by measuring faculty productivity including number of articles published (research and non-research) and number of citations in the literature. Their use of multiple criteria ranked departments rather than programs like science education. It should be noted that some institutions have a department of science education, other have a department of science and mathematics education; others are in departments of curriculum and instruction or other configurations. Perception studies seem to be influenced by visibility of the department and its members. Productivity Research Varied objective measures of institutional productivity have been applied over the years. The field of counseling psychology has historically ranked institutions based upon faculty members’ research productivity (Bohn 1966; Goodstein 1963; Howard 1983; Katz and Brophy 1975; Tinsley and Tinsley 1979; Delgado and Howard 1994). Initially, only the Journal of Counseling Psychology was used to rank programs (Bohn 1966; Cox and Catt 1977). Subsequently, several research journals in counseling psychology were used by Howard (1983) and Delgado and Howard (1994). Howard’s study focused upon journals published between the years 1976 and 1982 while Delgado and Howard used a ten-year period (1983–1992). The later study documented substantial changes in institutional rankings from earlier study (Howard 1983). Smith et al. (2003) used five education psychology journals from 1997 to 2001 to identify their top 35 programs. This was an extension of their study of educational psychology from 1991 to 1996 (Smith et al. 1998). There was considerable variation in institutional ranking of the two studies. There has been greater stability of rankings in industrial organizational psychology (Howard et al. 1985). Quality of programs has been based upon the research publications of faculty in mathematics (Davis et al. 1999), dentistry (Harrington and Lavine 1986), finance (Corrado and Ferris 1997) and economics (Scott and Mitias 1996). In addition to frequency of articles published by institutions’ faculties researchers have compared programs by the subsequent contributions to the field by their graduates. Several researchers (Bodenhorn 1997; Hartley and Robinson 1997; Schamauder et al. 1999; Robinson et al. 2001) have reported on research at liberal arts institutions. Bodenhorn and Hartley and Robinson’s research focused upon rankings for economic departments, while Schamauder, Robinson and Hartley’s research focused upon departments of psychology. Robinson, Hartley, and Dunn used geology publications in Geo Ref listed journals to establish their ranking. They concluded that geology publications are concentrated in a few institutions, but the number has increased since the 1970s by five folds. Notaro et al. (2000) ranked doctoral programs in health education based on faculty productivity and scholarly activity of doctoral students. They ranked 28 programs for each of eight variables (publications, citations received, editorship, external funding, student activity, student/faculty ratio, mentoring and placement, and student support) as well as overall ranking. They reported that the four faculty variables were not correlated with student variables. Earlier, Tauer and Tauer (1984) ranked agricultural economics programs based upon journal publications of recent graduates. This approach seems as potentially fruitful and promises to be an interesting line of inquiry. In addition to frequency of articles published by institutions’ faculty researchers have compared programs by the subsequent contributions to the field by their graduates. However, for this study, we elected to focus solely upon the productivity of the authors at institutions and not their graduates. Methodology The eight journals selected for this study represented a broad orientation of research in the discipline of science education (Yager 1984). Each of the journals was judged to have a low acceptance rate (\41%). These journals were selected because they predominately publish empirical research on science education and are common outlets for science education researchers working at domestic institutions of higher education. Journals which were not reported by Cabell (2002) were sent a personal communication (Table 1). Most journal editors were faculty at the major research institutions. Generally, there has been more than one editor for each journal for the duration of this study. Each of these journals focuses upon empirical research on science teaching and/or learning. Each of the journals has an editorial board which judges the merit of the submissions. Some individuals are reviewers for more than one of the journals. All journals except Science Education, International Journal of Science Education, and Journal of Science Education and Technology are associated with a professional organization. Our analysis included every issue of eight science education research journals from 1990 to 1999. Table 2 contains a listing of volumes, years, and institutional application for each of the eight journals. The author(s) and their institutional affiliation were reported for each article. It is acknowledged that not all authors have a primary interest in science education. However, their contributions are credited to their home institutions. The multiple journals perspective was recommended by Howard (1983). Since this is the initial productive study in science education, Delgardo and Howard’s (1994) recommendation of a ten-year period was followed. The review tried to use time frame where minimum number of editors were involved with each journal. We examined the population of each journal. We choose to do the population rather than randomly selecting articles. Every article that was published was included in the analysis except editorials, letters to the editor, and book reviews. For each article, a data form included title of article, researcher(s) name(s), and institutional affiliation. If more than one author was from the same institution, each was counted regardless of the department. Faculty who had moved during the 1990’s contributions were noted at the time of their submission. We employed two different approaches to calculate institutional productivity. We choose to give all researchers equal weight and a weighting value based upon order of authorship. We gave equal credit for faculty and graduate students since their combined efforts are indicative of the total program. This unweighted method relied upon raw counts. This approach has been used in previous productivity studies (Davis et al. 1999; Harrington and Lavine 1986; Corrado and Ferris 1997; Scott and Mitias 1996). The use of Howard et al.’s (1987) weighted, proportional counting system was also used to assign amount of credit for each article. Each article has a set value of 1.0. The formula: Individual Credit = ð1; 5n:1Þ=ðPn1 ¼ 11; 5i¼1Þp:218 where n = total number of authors and i = the individual author’s ordinal position. Possible combinations and resulting weighting for each co-author appears in Table 3. Total weighted values for each of the 30 programs was computed and subsequently ranked from high to low. This method focused upon the number of authors and position within the list that each contributor’s name appeared. Thereby, giving greater credence to the tradition where senior author has made the most substantive contribution to the manuscript. This study also differs from Robinson et al. (2001) who only credited institutions of first authors. Results There were a total of 374 different institutions that had a total of 2,547 research publications in the eight selected science education journals during the 1990s. There were 132 institutions (35.29%), which had only one science education research publication in the eight research journals. There was a wide variation in the number of institutions with research publications for specific journals. The result for each journal was: Journal of Research in Science Teaching (55.35%), School Science and Mathematics (43.32%), Science Education (39.84%), Journal of Science Teacher Education (30.75%), Journal of Science Education and Technology (22.46%), Journal of Computers in Mathematics & Science Teaching (15.24%), Journal of Elementary Science Education (14.71%) and International Journal of Science Education (13.90%) (Table 4). Using the total number of research publications by institutions (Table 5), a ranking of institutions was possible. The highest total number of publications was 103. There was a natural break between the 30th and 31st ranked institutions. The total for the top productive institution was close to five times the 30th institution, so there is considerable variation between institutions. Table 6 contains weighted publications by institutions. The total and weighted distributions are presented in Table 7 for institutions; the top three institutions (Purdue University, University of Georgia, and University of Iowa) were identical. For the top 10, there was a 90% agreement. The bottom 10 had a 70% agreement. Using the weighted value, the top ranked institution had almost six times higher value than the 30th ranked institution. Each journal was totaled for the top 30 programs and compared to the total number of publications in that journal for the 1990s. Science Education had the highest frequency with 55.84% of science education researchers from the top 30 institution. Other journals and frequency of top 30 researchers were: Journal of Research in Science Teaching (54.09%), Journal of Computers in Math and Science Teaching (51.72%), International Journal of Science Education (50.85%), Journal of Science Teacher Education (40.69%), Journal of Elementary Science Education (34.75%), School Science and Mathematics (32.32%) and Journal of Science Education and Technology (23.73%). Using the total, the Journal of Research in Science Teaching was the only journal where a research publication was from all of the top 30 science education programs. Science Education had publications from 25 of the 30, School Science and Mathematics with 24 of 30, Journals of Science Teacher Education with 23 of 30, and Journal of Science Education and Technology had 22 of 30. While International Journal of Science Education, Journal of Computers in Mathematics and Science Teaching and Journal of Elementary Science Education had 17, 14, and 12 respective of the top 30 programs. The rankings were compared for each journal of the top 10 institutions. For Journal of Research in Science Teaching, 4 of the top 10 programs have a lower ranking while School Science and Mathematics have five programs and two with zero publications from the top 10 programs. Science Education has 6 of the top 10 programs with a publication and four institutions with no publications. In contrast, only 2 of the 10 programs had similar rankings for publications in Journal of Computers in Mathematics and Science Teaching. While International Journal of Science Education had 4 of the top 10 programs represented. The Journal of Science Teacher Education and Journal of Elementary Science Education each had 5 of the top 10 programs with research publications. The newest journal, Journal of Science Education and Technology had only 2 of the top 10 programs represented. Discussion This study is the first effort in the field of science education to rank institutional programs based upon research productivity. According to Notaro et al. (2000), rankings are an effective mechanism for comparing institutions for any subject area. Regardless whether choosing unweighted or weighted ranking, programs can be compared with similar ranking especially for the top 10 science education institutions. This study provides campus administrators with a ranking of science education programs; thereby, providing another source of data when making decisions about the quality of their science education program. However, administrators should not take these rankings as definitive evidence of best science education programs. Like Smith et al. (2003), these findings are only the extent of publications in these science education research journals. We consider the low acceptance rate that these research journals are rigorous and represent diverse interests in the field of science education. Restricting the study to these eight journals would not capture the total institutional productivity of the science education programs. Barrow et al. (2002) found most prolific science education researchers published their work in other research journals (besides the eight used in this study) and non-research publications. Using a decade of research will identify more stable programs. The movement or retirement of a highly productive faculty can alter subsequent rankings. The field of science education has a wide variety of journals in which research can be submitted. Three of the journals seem to have a lower submission priority Journal of Elementary Science Education, Journals of Computers in Math and Science Teaching, and International Journal of Science Education. It is possible that faculty of some of the top 30 science education programs have a secondary focus, a non-technology orientation, and a non-quantitative orientation. It is also possible that certain journals (e.g. Journal of Research in Science Teaching, and Science Education) are given higher status in promotion and tenure or merit pay decision at the top 30 institutions. It is possible that some of the eight science education research journals are not available in the library or faculty’s personal library. The lack of availability could limit choices of where to submit research. White (2001) in his review of research publications noted the revolution of style since 1980. Specifically, he noted major changes in questions and topics (focus of the investigation), and subjects of the research. He reported increased interest in studies of learning strategies and metocognition. Since 1985, White reported there had been a decrease in the total number of articles published in the Journal of Research in Science Teaching and Science Education, but this pattern was reversed in the International Journal of Science Education. He noted the greater use of qualitative methodologies where manuscripts require more journal space to present the researcher’s position was occurring in Journal of Research in Science Teaching. White noted that some journals have a greater preponderance for quantitative research (e.g., International Journal of Science Education). This research has benefits for several reasons as noted by Robinson et al. (2001). It provides a benchmark to which other science education programs can compare their level of productivity. It can also be helpful in evaluating the research productivity of junior faculty. New science education Ph.D.’s who are seeking employment will see what the research emphasis at various institutions of higher education was in the 1990s. In addition to pragmatic reasons (e.g. tuition assistance, family, etc) graduate students can see how various programs and their faculty interests compared for research productivity in selecting which institution to pursue for their Ph.D. In addition, new science education faculty will be able to help establish personal realistic expectations for research (Boice 1992). Due to changing emphasis as described by White (2001), future studies could be conducted for a shorter period of time, i.e. five years. This is only meaningful after a longer term baseline study like this study. A pattern of types of research could be investigated to see if topics, designs, and related activities would vary for various science education research journals. These results provides a ranking of institutional productivity, a similar study could be conducted to identify the most prolific science education researchers for the 1990s. Implications From the position that retrospection provides, we can consider the implications of this data in preparation for potential follow-up studies of contemporary science education research. A reasonable approach of Table 7 would be to select the upper quartile or so of institutions and consider whether those were the most authoritative spokespersons for science education during the 1990s. Certainly in terms of research productivity, these institutions made the best showing. In addition, those institutions at the time employed noteworthy science education researchers. Left indeterminate in this data are questions about the influence of this research, the quality of the doctoral students produced by these institutions, or whether these institutions’ standings have persisted despite staffing changes through relocations, hirings, and/or retirements. But we do perceive that this data does provide grist for several mills. Several qualifying statements must be appending to these findings. First, we deliberately took a very narrow slice of the publications universe by focusing only on U.S. institutional affiliations, restricting the scope to a single decade, and using only science education research journals. An international perspective, a different time span, and a broader selection of journals could produce different results. Second, we made no effort to weight the relative influence of the journals. One could make the argument that certain journals should be assigned greater weight because these journals are more highly cited and thus have greater influence within the education research community. Third, the data were influenced by the demographics of the institutions at this particular point in time. Larger number of science education faculty, the age and tenure structure of the science education faculty, and their perceived value of published research by their local academic environment would likely influence these results. Another caveat is whether this data was indicative of overall education program quality. Certainly productivity in eight science education journals should not be used as the sole criterion for ranking institutions. For all of it’s flaws, the U.S. News & World Report rankings are based upon a constellation of criteria. Individuals seeking information about the relative quality of science education doctoral programs would be wise to consider other factors with Jablon’s 2002 report as one source of additional information for aspiring doctoral candidates to consider. The science education community represents a changing landscape. Such vagaries as changing administrative missions, hiring patterns, and retirement cycles would be expected to influence the composition of an institution’s science education faculty and hence the research productivity. As a result, institutions that may be ranked highly in one era might not fare as well in a different time period. Also, the actual number of articles published within a given volume may vary. In summary, these changes over time would suggest the potential value in comparing productivity, as measured through the weighted author method, across time periods and perhaps using intervals smaller than the decade that we used for our analyses. We would not suggest this single study is in any way definite. As a benchmark study, the work reported here provides a baseline for future studies and has practical benefits for other researchers (Robinson et al. 2001) as well as providing newer science education researchers realistic expectations for their research productivity (Boice 1992). We envision several subsequent avenues of inquiry. Hix (2004) compared departments by calculating the relative influence of authors by the number of subsequent citations of an original piece of work. Since the data we reported made no effort to chart ensuing influence on the community, this seems like a valuable consideration. In another intriguing approach, McCormick and Rice (2001) chose to evaluate institutions by comparing the productivity of the graduates of their programs; although, both of these studies focused upon political science departments there would seem legitimate reasons to explore the application to science education.

We will not be able to generate sufficient resources to preserve our university . . . if we do not find ways to defend the importance of both our public and our intellectual mission. We must not be preoccupied by internal quarrels or lose our resolve. We must find new ways to tell our story, while exemplifying the extent to which a public institution still inspires trust as well as commitment. – Chancellor Nicholas B. Dirks, University of California-Berkeley, Inaugural Address (2013) To what extent are universities displaying transformation in their core values? The ascendancy of market-like and market behaviours within and across public research universities has been well documented worldwide (e.g., Cantwell & Kauppinen, 2014; Slaughter & Rhoades, 2004). In their market-like efforts, these universities engage in competitions for external sources of money from research funding, industry partnerships, professors’ entrepreneurship, and student tuition and fees. As part of their market initiatives, public research universities seek to generate profits from activities including patenting, licensing, and other partnerships and contracts that could open revenue streams (Slaughter & Leslie, 1997). Focusing on the nature of such changes, the research literature suggests some fundamental reorientations of universities from serving social functions to profit motives. Yet some universities appear to show more blending in their organising principles than might be implied by this view. In recent years, research has questioned the extent to which public research universities have transformed themselves into industry-like, market-focused organisations (e.g., Bozeman & Boardman, 2013). Studies of student entrepreneurship, for instance, suggest that state-subsidised undergraduates leverage institutional and regional networks for social and environmental benefit (Mars & Rhoades, 2012). Such a finding, Mars and Rhoades conclude, indicates a ‘narrow organizational space that is a hybrid of two otherwise competing academic capitalist and public good knowledge/learning regime[s] . . ..’ (p. 453). Our opening quotation from Chancellor Dirks of the University of California-Berkeley suggests that the overarching orientation of higher education institutions may similarly encompass a level of hybridity. In Dirks’ vision, ‘public’ and ‘intellectual’ missions together form part of a single ‘story’ of the institution. To advance further a conceptual understanding of this organisational arena, the literature on institutional logics – ‘material practices and symbolic constructions’ that constitute a field’s organising principles (Friedland & Alford, 1991, p. 248) – shows promising potential for application to higher education. Theory on logics has been used to examine topics as diverse as the higher education publishing industry (Thornton & Ocasio, 2008), mutual funds (Lounsbury, 2007), healthcare (Reay & Hinings, 2009), science (Berman, 2012), and postsecondary governance (Bastedo, 2009). While there has been some conceptualisation of institutional logics that underpin higher education (Gumport, 2000), Gumport’s (2002) case study of Stony Brook University, the University of California-Berkeley, and the University of Illinois-Chicago (UIC) presents an all-too-rare empirical analysis of institutional logics at the campus level. Strikingly, the complexity of relationships among multiple logics remains less well understood than the ascendancy of one over the other (e.g., Slaughter & Rhoades, 2004). We aim to contribute conceptually and practically to understanding the interaction among institutional logics within US public research universities. While we focus on the US, building directly on Gumport’s 2002 case studies over a more recent time frame, our findings from that country, rooted in a tertiary system of strong governmental decentralisation and market coordination, may also prove illuminating in respect of global trends in this arena. Rizvi (2006, p. 66) observes the ‘direction of change’ in higher education policies worldwide to be ‘remarkably similar’, and signs of resistance and tension are equally global. To cite by way of example the issue of university marketing, we see the relevance of our findings to experiences of both Nordic and Australian universities. There, recent research has shown how the confluence of different value systems has brought about conflict and compromise (Onsman, 2008; Sataøen, 2015). Expanding on Gumport (2002), we draw on documentary materials to delineate the organising principles around which campus officials construct their articulation of university relevance. A cross-case comparative approach helps us tease apart variation in how each university may elaborate its own logic. Our study is thus situated within calls in the literature for theoretical development (e.g., Mars & Rhoades, 2012) and empirical analysis of how higher education leaders and managers express organisational positions on institutional relevance (Gumport, 2002). We assume that public research universities are heterogeneous organisations whose various logics may not entirely be known or available to all campus constituents. The operationalisation of logics on campus, a subject beyond the scope of this paper, may very well reveal its own form of tension and hybridity. Nevertheless, it is not entirely unreasonable to assign particular import to those logics which campus officials present externally, for these constitute a predominant organisational perspective. Background Institutional logics are the ‘belief systems and associated practices that predominate in an organizational field’ (Scott, Ruef, Mendel, & Caronna, 2000, p. 170) and as such can serve as a ‘template for action’ for organisations in the field (Bastedo, 2009, p. 211). The value of logics lies in the contention that the ‘pattern of an organizational design is a function of an underlying interpretive scheme, or set of beliefs or values’ (Greenwood & Hinings, 1993, p. 1055). Institutional logics, with attention to cultural-cognitive dimensions of fields and organisations, address this relationship (Gumport, 2000). As we have noted, many public research universities have increasingly embraced market-like and market behaviours. A core concern is the extent to which publicly subsidised institutions have fundamentally shifted from using these resources to serve the public good to advancing self-interest in making money. In this section, we draw on logics to explore ascendancy, coexistence, competition, and hybridity among the core perspectives that underpin public research universities. Researchers and analysts have recently focused on ways in which the field and policy environment constrains the logics available to organisations and individual actors. A field has been defined as ‘organizations that, in the aggregate, constitute a recognized area of institutional life: key suppliers, resource and product consumers, regulatory agencies, and other organizations that produce similar services or products’ (DiMaggio & Powell, 1983, p. 148). At the policy level, federal and state governments have increasingly incentivised the economic roles of research universities (e.g., Berman, 2012; Warshaw & Hearn, 2014). The targeted allocation of funding for research and development and workforce training in science and technology has been viewed as a primary trigger of academic capitalism (Cantwell & Kauppinen, 2014; Slaughter & Rhoades, 2004). Additionally, the loosened legal and regulatory environment, which broadens the scope and facilitates the ownership of patenting and intellectual property rights, also spreads market-mindedness among many institutions (Geiger & Sá, 2008). The field and policy environment does not control or predetermine organisational responses. But in behaviour and structure at least some universities appear to have repositioned research, teaching, and service around the interests of external stakeholders and markets. For example, institutional patenting of academic research was at first contested (i.e., pre-legitimate) but, following incentivisation by federal research policy and the rising biotechnology industry, became seen as potentially lucrative, normalised (i.e., legitimate) work for faculty, administrators, and campuses (Colyvas & Powell, 2006). Since the 1980s, within the context of state and federal economic competitiveness campaigns, US universities have increasingly articulated the economic value of their educational contributions to workforce and human capital development. Indeed, the community engagement activities of universities, ranging from student and faculty projects to consulting and training for local governments and leaders, are increasingly framed and categorised as serving economic development (Gais & Wright, 2012). Claims of economic relevance could suggest shifts in rhetoric rather than core organising principles and templates for action (Kleinman, Habinek, & Vallas, 2011). Slaughter and Rhoades’ (2004) observation thatmost universities do not makemoney fromtheir moves to the market may, on the one hand, suggest institutional adaptation to emulative/normative pressures within the field of higher education. But, on the other, foundational studies in this arena do detect coexistence of underlying logics that has hitherto been largely overlooked. As Slaughter and Rhoades acknowledge, ‘academic capitalismhas not replaced the public good knowledge regime. The two coexist, intersect, and overlap’ (p. 29, emphasis added). We see recent literature on logics as especially helpful because it focuses less on ascendancy and more on examples of entwinement of market efforts and commitment to the public good. We are particularly struck by Mars and Rhoades’ (2012, p. 455) caution against regarding academics and units within ‘the capitalist domain of the academy’ as being ‘disconnected entirely from the public good underpinnings of higher education’. Recognition that a new logic need not supplant an existing one within an organisation, but might equally coexist with it, has led to recent discussion of the hybrid organisational forms adopted to accommodate coexistence (Greenwood, Raynard, Kodeih, Micelotta, & Lounsbury, 2011; Mair, Mayer, & Lutz, 2015). At the same time, emphases on coexistence may belie contestation and competition. As Clark Kerr (1995) has observed, the public research university has become so many things to so many different people that it must out of necessity be at war with itself. Analyses of coexistence may not quite detect ‘under the radar’ logics, which can emerge and take root when opportunities in the external environment open pathways for organisational change (Reay & Hinings, 2009). The use of institutional logics to understand change and adaptation – organisational responses – of public research universities has remained somewhat neglected. Gumport’s 2002 study found that ‘social institution’ logic became ‘industry’ logic, motivating academic restructuring for short-term profits and positioning around the funding interests of external stakeholders. One institution, however, evinced market goals but affirmed in structure a ‘social institution’ logic, which to Gumport indicated a divergence between rhetoric and practice. Such complex interactions, she acknowledges, remain only partially understood. It is to explore this complexity that we have undertaken the current study. We acknowledge the importance of examining triggers of market-like and market behaviours of public research universities, such as federal research policy, state finance and governance, and competition in the field of higher education. Yet our focus here is organisational responses as observed through relationships among logics within and across public research universities. To this end, we ask: (1) In what terms are the predominant organising principles of public research universities expressed? (2) What does this tell us about relationships among institutional logics, within and across public research universities over time? Research design and method Following Gumport (2002), our study data come from Stony Brook University, University of California-Berkeley, and UIC. The three are, Gumport concludes, appropriate for comparative analysis. All are public universities classified as ‘R1: Doctoral Universities (highest research activity)’ under the current Carnegie basic classification. Each experienced a marked decline in the share of its revenue coming from the state between 1980 and 1994. Despite some variation in this indicator, the trend continues downward (College Board, 2015). They face a somewhat comparable state-funding climate, a context of persistent ‘resource turbulence’ (Gumport, 2002, p. 57) and also, as research institutions, similar influences from federal research policy and resource competition. Similarities notwithstanding, the three are by no means entirely comparable. Only two – Stony Brook and Berkeley – are members of the Association of American Universities for leading North American research universities. In size, Stony Brook and UIC are most alike with approximately 20–25,000 full-time equivalent students to Berkeley’s 35,000. Both are also relatively young, UIC tracing its foundations to 1946 and Stony Brook to 1957; by contrast, Berkeley is nearly 150 years old. While we have examined these universities partly to develop Gumport’s 2002 analysis and partly because we too observe fundamental similarities between them, such differences are themselves of interest. They have been founded at different times for different purposes and are located in different environments, evolving as organisations – in structure, operations, ambitions, and goals – in distinct ways. There is reason to believe, then, that each might articulate its core principles differently, so enriching our understanding of the interplay of logics. We contend that the terms in which universities present their missions tell us something important about underlying institutional logics. To this end, our research explores the written record. Our evidence comprises publicly available documents from the universities’ websites. These covered the period 2000–2014 and included strategic plans, accreditation self-studies, and commissioned reports (see Tables A1–A3 in the Appendix). In addition to key mission and planning documents, we analysed all records in which campus leaders expressed a position on institutional relevance. Press releases and other short statements lacking significant contextualisation were discounted. Transcripts of speeches made by senior officers, addressing a range of audiences, were included. While we are well aware that institutional practice may diverge from written policy, we maintain that the written record serves an important purpose. Despite criticism that such documents are simply ‘rhetorical pyrotechnics’ – ‘pretty to look at perhaps, but of little structural consequence’ – they play a complex yet important part in communicating underlying values (Morphew & Hartley, 2006, p. 456). We argue that strategic plans and self-studies have a particularly important role as generative documents (Prior, 2003), setting the boundaries for discussion of university form and purpose. What the universities choose to focus on points us towards the sources from which they seek to draw legitimacy. We employed a mixed strategy for cross-case analysis (Miles, Huberman, & Saldaña, 2013). A first cycle of descriptive coding, undertaken in NVivo (QSR International, Doncaster, Victoria, Australia), coded all passages that mentioned one or more of the universities’ three core missions – research, teaching, and engagement. This enabled us to focus more closely on key passages from our data. Some 1,200 pages were subject to first-cycle coding, from which passages totalling approximately 200 pages were extracted under one or more of the three themes. Our second-cycle coding was in part elaborative, taking account of the ‘theoretical constructs’ (Saldaña, 2013, p. 229) of Gumport’s (2002) findings. Our aim was not to recreate the previous research, but to probe the themes which emerged from it. Gumport’s research therefore served as a point of reference. The substantive focus of our coding, meanwhile, was the logics that were being expressed. While some are explicitly stated, many more are implicit in what is said. To capture both explicit and implicit value statements, we recorded what research, teaching or engagement activity was being discussed, what rationale or purpose was ascribed to it, and who specifically was being implicated (both within the university and externally). Repeated crossreferencing enabled us to identify tensions between statements, make connections between documents, and identify change or consistency over time. These elements were drawn together in a series of memos, one per institution for each theme, from which the cross-case analysis presented here emerged. Our decision to focus solely on documentary evidence made available by the institutions occasions a note of caution. A decisive pronouncement on the mix of logics drawn on and operationalised by each university would require reference to a more comprehensive range of sources – interviews, observations, internal memoranda, and the like. It would also demand investigation of the interaction between university and field, including the university system within which each operates. Still, we see our study as building our understanding of the interplay of logics. Although we do not trace the (potentially conflictual) processes that have shaped these publicly available documents, as the outcome of those processes these documents present viewpoints that carry particular authority. Findings Our initial investigations identified the three core missions of research universities – research, teaching, and engagement with external communities – as the intersecting cross-case themes around which further analysis could best be structured. In answer to our first research question, we present our findings on each in turn below. Having done so, the following section draws these together in discussion of the relationships between different logics within and across our study institutions. Research For all three universities, research was inevitably of prime importance, what Berkeley’s 2002 strategic plan called the ‘energy’ that fuels the University’s mission. But beyond a shared commitment to research, how did each university express its research mission? At his 2009 inauguration, Stony Brook’s President Stanley referenced the institution’s founding mission of ‘excellence in science and technology’, and this legacy has remained strong. Scientific and technological disciplines received notable attention in the documents studied. Partnership with the Brookhaven National Laboratory, for instance, appeared as a priority as early as the 2000–2005 strategic plan, reappeared in the 2008–2013 plan, and was highlighted as recently as the 2013 state of the university address. At Berkeley, meanwhile, ‘breadth and depth’ of research disciplines was emphasised. Although 2012 saw publication of two reports detailing a narrower vision of Berkeley’s core disciplines, by 2013 Chancellor Dirks was using his inauguration speech to ‘resist the stark divide between . . . basic and applied research, between the arts and the sciences . . .,’ and to once again champion excellence and innovation across the broadly defined scope of Berkeley’s activities. In discussing its research mission, UIC, rather more explicitly than either other institution, tied what is being researched to why. Like both Berkeley and Stony Brook, UIC championed interdisciplinarity. It also, as befits an institution with a long-standing commitment to the medical field, particularly emphasised health research. But most strikingly, in its 2006 strategic plan it described the contribution which UIC could make to ‘new knowledge . . . that produces unique perspectives, solutions and understanding of our lives, society and the natural world’. This is not to say that the audience for university-generated knowledge was deemed to exist solely outside academia. All three universities repeatedly emphasised elements of what we might term a more purely ‘academic’ value set. Both UIC and Stony Brook identified their researchers as ‘pioneers’ operating at the ‘frontiers’ of research. This imagery found ultimate expression at Berkeley where at his inauguration Chancellor Birgeneau cited Chancellor Emeritus Seaborg thus: The spirit of our pioneering past is the spirit we must seek for our present and future . . . . Learning and discovery are the New Worlds and the Old West, the lands of opportunity. The pursuit of research excellence is a particularly interesting commitment since it ties to the institutions’ competitive context. The universities all stated their aspirations and achievements in terms of increased national and international standing. Nevertheless, there remained a degree of ambivalence towards such competition. Referring to the latest university rankings in his 2012 state of the university address, Stony Brook’s President Stanley succeeded in simultaneously appealing to and criticising them: The recent U.S. News & World Report rankings placed us in the top 100 national universities and among the top 40 public universities – the highest ranking we have ever achieved. And while I am always skeptical of these ratings, I hope they reflect the real progress we have made. Yet if rankings are a source of scepticism, why pay them attention? In part, the institutions’ own mission statements answer this question. Excellence in research appeared as a goal in its own right, and league table success as evidence of its successful pursuit. There was, however, a potential subtext, namely that evidencing excellence aids the quest for funding. There are clear commonalities in the universities’ discussion of their research missions. Nonetheless, there are also differences in the degree to which they tended to promote one rationale over another. The first spectrum on which the institutions sit relates to who research is deemed to serve and how. Fairly firmly at the ‘scholarship’ end of the spectrum sits Berkeley, with Stony Brook tending more decisively to the ‘engagement’ end. While UIC noted that all research has a context, it sits somewhere between, embracing both academic and external communities. A second spectrum represents the extent of breadth or focus in the research mission. All three highlighted interdisciplinarity and its importance to excellent research. However, the founding principles of each led to somewhat different emphases: liberal arts at Berkeley, science and technology at Stony Brook, and healthcare and urban studies at UIC. Students and teaching As Stony Brook affirmed in its most recent accreditation self-study, ‘our students are at the core of what we are as a university, and why we do what we do’. To understand how each university articulates its teaching mission, three questions prove helpful. Who does the university say should be taught? What are they being taught? And why are they being taught those things? The available documentary evidence suggests a recent shift in the universities’ understanding of who should constitute the student body. In 2007, for example, Berkeley highlighted the fact that 90% of its undergraduates were from in-state. By 2011, the target was to raise the out-of-state and international undergraduate population to 20% of the student body. ‘We cannot’, argued Birgeneau, ‘sustainably afford to educate significant numbers of Californians for whom we receive no state funding . . .’. Stony Brook has similarly implemented a target of 30% for out-of-state and international students. Attracting students from outside the region, it noted, generates additional spending in the University and the wider economy. Interestingly, however, a year after the strategic plan pledged that in-state student numbers would fall to 70% of the total, President Stanley’s inaugural speech highlighted the fact that 85% of students were currently from in-state – in other words, that ‘we are educating New Yorkers’. This observation came in the context of a speech that described Stony Brook as ‘an absolutely vital part’ of New York’s economic recovery and made an emphatic case for increasing state funding of Stony Brook. Wherever they may come from, what is it that students are being taught? Here, Berkeley stood apart from UIC and Stony Brook with a more explicit commitment to the primacy of a liberal arts education. UIC, in contrast, positioned itself predominantly as providing education that prepared students ‘for the world in which they will be citizens’. Where UIC made repeated reference to professional education, Stony Brook identified expansion in this area alongside creation of a ‘core curriculum’ for undergraduates that encourages critical thinking and ‘basic familiarity with the power of science, technology, the arts, humanities and social sciences’. The relationship between breadth and purpose of the education mission is not a straightforward one. At the same time as drawing attention to the broad-based nature of what it teaches, in respect of outcomes Stony Brook focused clearly and repeatedly on the economic and workforce benefits of higher education. UIC, in an inversion of this pattern, placed professional education programmes far more centrally in its discussion of what was being taught, but defined the outcomes and beneficiaries of this education more broadly. The relationship might have been held at Berkeley, but elsewhere breadth of taught content – or at least the approximation of this offered by provision of a liberal arts education – and breadth of anticipated outcomes did not go hand in hand. As to why the universities provide the education they do to the students whom they enrol, two related yet distinct forms of benefit were noted. Each university identified both forms, and so the distinction between them was a matter of degree. The first, broader interpretation of benefit envisaged positive social outcomes from higher education that included graduates who are willing and able to serve the public good as more productive, engaged citizens and better leaders. The second interpretation focused more exclusively on the private benefits of higher education. These included both individual benefits – notably improved employment prospects – and the advantages for the American economy attendant on this contribution to the workforce. Berkeley stressed service to the public good more overtly and more frequently than did either other university, and in 2014 Chancellor Dirks expressed concern that ‘education is increasingly seen as a private good at best’. Both Stony Brook and UIC made repeated reference to their role of meeting workforce needs. Nevertheless, the social good – and particularly the need to instil the values of citizenship – remained a theme throughout the study period. In addition, there was some evidence that public and private benefits were regarded as complementary, as when Stony Brook identified success in placing graduates in the job market as evidence of its relevance to social needs. Similarly, improving access was promoted for both its public benefits (improving social justice) and its private (largely economic) ends. Engagement with external communities Engagement with external communities formed the third part of the institutions’ stated missions. What form of engagement was emphasised varied not only between institutions but according to the audience for a particular speech or report. The acknowledged benefits of engagement can be divided broadly into three categories. The first includes those accruing to industry, to the workforce, and to the economy more broadly. The second concentrates on quality of life and the benefits of engagement activities to society at large, with healthcare a particular focus of attention. Thirdly, the university itself was seen as a beneficiary, through engagement’s impact on research, on teaching, on university reputation, and on income flows. In the documents studied, the intersection of these varied outcomes with different audiences and scales of operation produced a complex picture. At Berkeley, the emphasis tended towards engagement for the benefit of society. Specifically, service to ‘the people of California’ has been consistently stressed. This service is both direct, encompassing the benefits of research, and indirect, arising from the education of ‘a new generation of leaders, innovators, and educators’. The University’s liberal arts focus was considered central to this outcome. Successive chancellors have described the ‘privilege’ of being at Berkeley and the ‘obligation to give back to society’ attendant on it, and they have committed the University to ensuring that students ‘ask themselves and each other the most challenging moral and intellectual questions about meaning and purpose in our lives and in our society’. Preparation of students for ‘public service’ and ‘leadership’ was central to Berkeley’s discussion of engagement. The economic mission, meanwhile, undoubtedly figured more strongly at Stony Brook than at Berkeley. While the latter acknowledges an economic function, emphasis was most pronounced between 2005 and 2012 and has since declined. The former’s rationales for engaging with third parties included acting as a ‘regional economic engine’, filling the gap left ‘as the private sector pulls out of the research arena’, and helping ‘to create student employment opportunities’. Yet the engagement mission even here remained tied to enhancing ‘quality of life’ more generally – including through contribution to the region’s health and a commitment to answering ‘the big questions, the questions that matter’. UIC’s engagement mission was subtly different again. Where Stony Brook referenced the region and state, UIC defined itself as ‘inextricably tied to the city of Chicago’. Through its Great Cities initiative, the University has for over 20 years ‘sought to improve the quality of life’ in its home city (and in other ‘great cities’ worldwide). And although UIC, too, highlighted the economy as a key to urban development, in its references to ‘job-centered development’ and direct engagement between the UIC Center for Urban Economic Development and ‘community organizations, labor unions, employers and government’, there was a stronger sense of commitment to the people of Chicago than to its economy per se. It is manifest, then, that private and/or economic goals were not pursued in isolation from public and/or social ones. Competition between universities was in evidence in the engagement mission as in research. In UIC’s 2013 strategic planning document, success in engagement was overtly linked to institutional reputation. At Berkeley, the 2008 vision document identified engagement activity as a source of competitive advantage: excellent performance coupled with accessibility to students from ‘across all socioeconomic levels’, it claimed, made the University ‘uniquely attractive to faculty who want to serve the public good’. This cannot, however, be taken to imply that the universities have complete control over the activities they pursue. Although discussion focused primarily on the universities’ chosen actions, certain phrases hinted that at least part of the engagement mission was externally driven. UIC’s strategic plan described how the growing expectation of economic development as a component of university missions places greater emphasis on the creation of new knowledge and the commercialization of innovations. Moreover, by focusing on the financial benefits to the institution of engagement activity, each university at times pointed to engagement being a means to an end rather than a core mission in its own right. This connection, though, remained largely unarticulated, and both Stony Brook and Berkeley also claimed to be incurring costs in fulfilment of their engagement missions. Discussion The findings presented above advance our understanding of the three universities’ application of institutional logics in a number of ways. Firstly, reinforcing Gumport’s 2002 findings, both ‘industry’ logic – in which appeals to legitimacy are made with reference to market forces, economic development and workforce skills – and ‘social institution’ logic – which draws legitimacy from social goals as well as ‘traditional academic ideals’ (p. 54) – continued to coexist. It would therefore seem that Gumport’s contention that ‘organizational discourse about goals and solutions came to be cast in an industry logic’ (p.73), that industry logics now predominate, cannot be wholly justified. In our study, we do not observe a wholesale undermining of educational missions – that key part of the social institution logic (e.g., Bozeman & Boardman, 2013). Moreover, Colyvas and Powell (2006, p. 315) have suggested that ‘the presence, absence, onset, and cessation of commentary can be utilized to periodize the development of an institutional rule or organizational form and to develop simple categorical measures of legitimacy’. On this basis, mention throughout the study period of values that correspond with a social institution logic – from an ‘academic’ value set in the research mission to the social justice outcomes of engagement – implies that this logic remains a valid one. In its unique balance of disciplines, each university displayed a distinct orientation. Gumport observed UIC’s health science focus, Berkeley’s comprehensive field coverage, and Stony Brook’s techno-scientific bent in 2002, and they remained apparent. Our data also highlighted persistent differences in the spatial scales with which the universities identify when undertaking engagement work. While Berkeley stressed its responsibility to the state of California, UIC defined itself as being ‘inextricably tied’ to its home city of Chicago. This calls into question the extent of institutional isomorphism across our study institutions (cf., DiMaggio & Powell, 1983). While the missions and values of the three universities have some similarities, they were emphatically not the same. Berkeley’s Chancellor Dirks proposed that only through continued adherence to longheld values will the University justify and support its existence. The professed importance of continuity of mission suggests that inter-university differences will persist. This is not to deny tension and even contradiction between the logics employed within each university. Our findings show evidence that, on occasion, there is just such contradiction in the universities’ positions. Industry logics were employed in the study institutions, and evidence pointed to institutional behaviour having altered in response to external pressures. Recent shifts in the balance between in-state and out-of-state students, for example, were justified on the grounds that declining state appropriations necessitated them. Yet the picture is far more complicated than simple replacement of social institution logic with industry logic. In 2008, Stony Brook pledged to increase the proportion of out-of-state students in a move that, it argued, would bring new private spending to the University and region; conversely, in 2009 President Stanley used the fact that in-state student numbers remained high to make the case for state funding. While we might infer from the 2008 target that industry logic is driving a new approach, this selective use of figures suggests that something more complex is occurring. A possible explanation is that we are observing a kind of system-gaming behaviour in which, when deemed expedient, symbolic adherence to one logic facilitates preservation of an identity tied to a separate logic (Greenwood et al., 2011). Identification with and adoption of particular logics is an inherently political process (Bastedo, 2009), and we cannot overlook the political capital to be gained by adopting particular positions as audience dictates. A diagnosis of symbolic adherence would certainly help to make sense of, for instance, the universities’ foregrounding sometimes of the rising proportion of out-of-state students – to demonstrate financial sustainability – and sometimes of continued high proportions from in-state – to demonstrate ongoing relevance to the state. System-gaming might also account for our observation that certain standpoints (that the state benefits economically from a highly educated population, for instance) were foregrounded even as others (such as the observation of growth in federal research funding levels) were relatively sidelined. Elements of our data, however, appeared to show mutual dependence between the logics, in which adoption of an industry logic gave renewed impetus to a social institution logic. UIC’s self-defined role as ‘regional economic engine’, for instance, was linked to its commitment to improving quality of life among Chicagoans. Without reference to their social benefits, we can only partially understand the University’s economic development goals. Social goals, meanwhile, relied at least in part on a strategy of ‘job-centred development’ for their realisation. Similarly, in evaluating its relevance to social needs, Stony Brook identified graduate job market success as key evidence. Berkeley’s commitment to improving ‘agricultural and industrial productivity’ was made in the context of service to ‘the people of California’. To overlook this connection would be to apply a partial reading that misrepresents the interconnectedness of the economic and social goals. Evidence from our study universities thus accords with Bastedo’s (2009) and Mair et al.’s (2015) identification of organisations that recognise the potential inherent in the combination of multiple logics. Whereas these studies describe the organisations as hybrids, with the logics remaining recognisably distinct, our own research points to hybridisation of the logics themselves. We argue that when elements of a university’s organising principles can only be understood in relation to the interaction between logics, those logics can no longer be regarded as truly separate. Concluding remarks Collectively our findings show multiple strategies at play in the universities’ responses to concurrent demands of social institution and industry logics. No single response can be observed, even within a single university. Of the responses we observed, we are most interested in hybridisation – the blending of industry and social institution logics to create a new form. Slaughter and Rhoades (2004, p. 336) conclude that: ‘Rather than simply seeking to maximize external revenue generation, academic capitalism could seek to enhance the social benefits of intellectual property and educational services’. In uncovering instances in which elements of an industry logic give renewed impetus to a social institution logic, we demonstrate (albeit in a limited way) that such an alternative is employed in some universities’ statements of institutional relevance. The concept of hybrid or blended logics suggests a promising framework for understanding how universities can and do manage and exploit tensions in their missions. We have thus far looked at the largely theoretical world of mission definition and see merit in exploring the hybridisation of logics in practice. What examples of hybridisation can be identified in the practices of these and other universities? How is the hybrid logic brought about? And, importantly, what are the limits: under what circumstances will industry logic support social institution logic, and under what circumstances supplant it? Sataøen (2015, p. 714) has called for research into ‘the process whereby core values . . . are negotiated and built’ so as to better understand how universities manage tensions in their mission. Further research into the role of hybrid logics would seem to answer that call. In addition to its potential descriptive power, we contend that such research could also make a valuable contribution to universities’ strategic planning processes. As universities seek to position themselves amid potentially contradictory or competing logics (Rizvi, 2006), answers to the questions posed above might allow us to better predict circumstances under which these logics could be aligned in a mutually beneficial, hybrid form. Appropriately applied, the findings could serve as another tool in universities’ multifaceted response to institutional complexity. Copyright of Journal of Higher Education Policy & Management is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Institutional research has been defined as ‘‘The sum total of all activities directed at describing the full spectrum of functions (educational, administrative and support) occurring within a college or university. Institutional research activities examine those functions in their broadest definitions, and embrace data collection and analytical strategies in support of decision-making at the institution’’ (Middaugh et al. 1994). As a discipline, institutional research stands at the nexus of collecting, analyzing, and providing consistent, reliable information. This information nexus has expanded to include not only objective facts about the institution, but also information about the extent to which students are achieving expected competencies in programs of study. The proportion of postsecondary institutions in the United States engaged in activities connected with the assessment of student learning has risen steadily over the past two decades, and now encompasses the vast majority of colleges and universities in the United States, including doctoral research institutions. Thus, to the above definition must be added the increasing responsibility faced by many institutional researchers for assuring the ongoing assessment of student learning, the reporting of assessment findings, and the usage of results across the institution for ongoing programmatic improvements.

The rise of accountability After nearly a half-century of institutional research in higher education, multiple data sources exist in our institutions. At the national level a main focus in the rise of accountability is our national reporting done through the Integrated Postsecondary Education Data System (IPEDS). In addition, the rise in accountability has been facilitated with the development of a web-based data collection system for IPEDS and web-based retrieval tools that allow the public to access IPEDS data. However, institutional reports have been designed primarily for internal managerial usage, and have not hitherto been consistently presented in a manner that facilitates stakeholder understanding of the extent to which their goals are being met. The existence of a well-educated population is primary to a nation’s economic development, and thus attention is increasingly devoted to driving up the proportion of a nation’s population that has attained a college or university degree (Campbell et al. 2007). Viewed in this light, the performance of a nation’s colleges and universities in producing well-educated graduates becomes a matter of national economic security and is bound to be closely scrutinized by those whose interests are at stake. With the increase in public demand for transparency and accountability for results, information from offices of institutional research is currently being re-sculpted into ways that are more consumable. Consumers and institutional stakeholders are now taking an active role in demanding this information and calling for institutions to provide data in a more transparent fashion. At the federal level in the United States, questions of institutional accountability for resources consumed and results achieved have garnered national attention (Spellings Commission 2006). Answering this, organizations such as NASULGC, AASCU, and NAICU have posted information templates for institutions to follow, and in the face of public pressure, hundreds of public and private institutions have elected to participate. The intent is to provide a simple, understandable framework to tell an institution’s story to those who are interested and to facilitate the decision to apply for admission to a particular college or university. The rise of global competition in higher education At the institutional level, as the global marketplace has flattened so too has the market for higher education become global (Friedman 2005). The need to continue a steady stream of qualified students in the face of declining birth rates in many developed nations, coupled with decreases in state-provided funding and a flattening environment for research funding have forced major research universities, liberal arts institutions, and specialized institutions to seek new sources of both students and revenue to continue and expand their activities. One outcome of this is greatly heightened competition for resources. American, British, Australian, and institutions from other nations now compete around the globe, wherever a market for higher education exists. The advent and exponential growth of the for-profit educational sector worldwide has further accelerated the competition for resources. Typical avenues of addressing the need for increased student and financial resources consist of an increased off-campus presence, including study abroad sites, satellite campuses, distance/elearning programs, and for many institutions, the formation of strategic partnerships, research alliances, and joint degree programs with similar institutions in other regions and countries. The emergence of academic analytics The need to demonstrate accountability to its stakeholders and the rise of global competition has brought to the forefront institutional research, a discipline rooted in the philosophy of fact-based management for progressive institutional improvement. As institutions become aware of the need to contribute to national economic development, compete in a global environment, understand and improve their internal processes, and provide consistently highquality student outcomes, conventions of standard and ad hoc reporting in institutional research have given way to an emphasis on statistical analysis, data mining, forecasting, predictive modeling, and optimization studies to inform decision making and increase competitive advantage, a development referred to as academic analytics (Goldstein 2005; Campbell et al. 2007). The body of this paper is concerned with the presentation of case studies that exemplify how institutional research has served to promote improved decision-making through the analytic lens it brings to bear, the decisions such analytic capabilities facilitate, and the institutional improvements that result. Case studies The use of institutional research in the private sector The first case study from the private sector will feature a mixed methodology study that used data from the National Survey of Student Engagement (NSSE) and student focus groups to explore student engagement at this institution and inform faculty development. In addition, an example of how institutional research informed decision making from the standpoint of a specialized private institution.

Springfield College Springfield College is a private master’s level institution with a unique mission. For more than a century, Springfield College has been a leader in providing a broad and balanced educational experience. Humanics, the philosophy that has inspired Springfield College (SC) from its beginning, calls for educating the whole person—spirit, mind, and body—for leadership in service to others. Student engagement informs faculty development Method Given its unique mission and history, as a community SC prides itself in high levels of faculty–student interaction and in providing active learning environment for all students. The institution supports the theoretical framework of the NSSE (Kuh 2003). NSSE is built off of the tenet that what students do during college counts! In addition, SC believes that its experiences are centered around Chickering and Gamson’s (1987) ‘‘Seven Principles for Good Practice in Undergraduate Education.’’ These principles include student–faculty contact, cooperation among students, active learning, prompt feedback, time on task, high expectations, and respect for diverse talents and ways of learning. As a result, it was counter intuitive to find results from the NSSE that indicated lower than average faculty– student interaction and lower benchmark engagement scores for Springfield compared to our comparison group. As a result of these negative trends in the quantitative data, institutional research in conjunction with the Outcomes Assessment Task Force (OATF) sponsored a focus group project to gain the insights of current students. The purpose of the study was to examine student perceptions gleaned from quantitative NSSE data that were not consonant with perceptions of the undergraduate educational experience at Springfield College. Current seniors were randomly selected to participate in the project and the sample was stratified to represent various academic programs. Students were provided an incentive to participate in the focus groups. Four focus groups totaling 30 participants ranged in size from six to nine seniors. Each group met for a 2-h session. Focus group questions were refined over a period of months, moving from the proposal draft, through input from the OATF, to minor revisions of the script after the first focus group session. Two key questions addressed the study’s purpose: how would you characterize your experiences with Springfield College classes and professors, and how would you summarize your academic experience? Themes were derived from the emphasis participants placed on particular aspects of their academic experience at the college. Themes represented experiences that were common across both individual participants and group sessions. More importantly, the themes were selected to provide potential insights about students’ (1) sense of academic engagement, (2) relationships with faculty, and (3) perceptions of the college experience that differed from those reported through the quantitative portion of the study. Particular examples and contrasting assertions provide glimpses of academically engaging classes, motivating relationships, and the college experience as seen with student eyes. Results Results of this qualitative project indicated that participants valued engagement with faculty, particularly when those relationships are embedded in the routines of campus life—classes, advising, and going about normal daily activities. Participants elaborated on their reasons for choosing the college, their transition to college-level academics, the evolution of their academic experiences over several years, and the types of classes that they found to be engaging. Some consistent themes that emerged included: Campus—Home Away from Home—Corny or Not, Advising—An Inconsistent Adjunct to the Academic Experience, All College Requirements—Academic Regrets and Broadened Horizons, A Teacher Makes a Difference— Caring and Challenge, The Big Picture—Bringing It All Together. In some cases these themes were thought to contradict the NSSE findings while in other cases they gave support and cause for pause for the institution. As a result, Institutional Research and the OATF felt that this data needed to be shared with the faculty. Thus, the results from both the quantitative and qualitative portions of this study were presented to the leadership of the Faculty Senate, who then organized a Faculty Institute to present faculty with the opportunity to learn about and discuss the NSSE findings and the focus group findings. In planning this institute, all involved parties, Institutional Research, OATF, and Faculty Senate worked collaboratively to plan specific workshops focused on the topic of student engagement in learning and used the very powerful voices of their own students to engage faculty in a review of their classroom procedures and interactions with students. This initial institute sparked a series of faculty development opportunities that were sponsored by the Faculty Senate over the next several academic years. As part of the institutional assessment plan, SC continued to participate in NSSE on an every other year cycle. Interestingly enough in the administration of last cycle of NSSE—2006, all five of the national benchmark scores for SC were higher than the corresponding benchmark. Savannah College of Art and Design Savannah College of Art and Design (SCAD) is a private master’s level institution focused on art, design and architecture, with campuses in Savannah and Atlanta, Georgia. The 2006 enrollment comprised over 8,900 students from 108 countries. Springfield College Springfield College is a private master’s level institution with a unique mission. For more than a century, Springfield College has been a leader in providing a broad and balanced educational experience. Humanics, the philosophy that has inspired Springfield College (SC) from its beginning, calls for educating the whole person—spirit, mind, and body—for leadership in service to others. Student engagement informs faculty development Method Given its unique mission and history, as a community SC prides itself in high levels of faculty–student interaction and in providing active learning environment for all students. The institution supports the theoretical framework of the NSSE (Kuh 2003). NSSE is built off of the tenet that what students do during college counts! In addition, SC believes that its experiences are centered around Chickering and Gamson’s (1987) ‘‘Seven Principles for Good Practice in Undergraduate Education.’’ These principles include student–faculty contact, cooperation among students, active learning, prompt feedback, time on task, high expectations, and respect for diverse talents and ways of learning. As a result, it was counter intuitive to find results from the NSSE that indicated lower than average faculty– student interaction and lower benchmark engagement scores for Springfield compared to our comparison group. As a result of these negative trends in the quantitative data, institutional research in conjunction with the Outcomes Assessment Task Force (OATF) sponsored a focus group project to gain the insights of current students. The purpose of the study was to examine student perceptions gleaned from quantitative NSSE data that were not consonant with perceptions of the undergraduate educational experience at Springfield College. Current seniors were randomly selected to participate in the project and the sample was stratified to represent various academic programs. Students were provided an incentive to participate in the focus groups. Four focus groups totaling 30 participants ranged in size from six to nine seniors. Each group met for a 2-h session. Focus group questions were refined over a period of months, moving from the proposal draft, through input from the OATF, to minor revisions of the script after the first focus group session. Two key questions addressed the study’s purpose: how would you characterize your experiences with Springfield College classes and professors, and how would you summarize your academic experience? Themes were derived from the emphasis participants placed on particular aspects of their academic experience at the college. Themes represented experiences that were common across both individual participants and group sessions. More importantly, the themes were selected to provide potential insights about students’ (1) sense of academic engagement, (2) relationships with faculty, and (3) perceptions of the college experience that differed from those reported through the quantitative portion of the study. Particular examples and contrasting assertions provide glimpses of academically engaging classes, motivating relationships, and the college experience as seen with student eyes. Results Results of this qualitative project indicated that participants valued engagement with faculty, particularly when those relationships are embedded in the routines of campus life—classes, advising, and going about normal daily activities. Participants elaborated on their reasons for choosing the college, their transition to college-level academics, the evolution of their academic experiences over several years, and the types of classes that they found to be engaging. Some consistent themes that emerged included: Campus—Home Away from Home—Corny or Not, Advising—An Inconsistent Adjunct to the Academic Experience, All College Requirements—Academic Regrets and Broadened Horizons, A Teacher Makes a Difference— Caring and Challenge, The Big Picture—Bringing It All Together. In some cases these themes were thought to contradict the NSSE findings while in other cases they gave support and cause for pause for the institution. As a result, Institutional Research and the OATF felt that this data needed to be shared with the faculty. Thus, the results from both the quantitative and qualitative portions of this study were presented to the leadership of the Faculty Senate, who then organized a Faculty Institute to present faculty with the opportunity to learn about and discuss the NSSE findings and the focus group findings. In planning this institute, all involved parties, Institutional Research, OATF, and Faculty Senate worked collaboratively to plan specific workshops focused on the topic of student engagement in learning and used the very powerful voices of their own students to engage faculty in a review of their classroom procedures and interactions with students. This initial institute sparked a series of faculty development opportunities that were sponsored by the Faculty Senate over the next several academic years. As part of the institutional assessment plan, SC continued to participate in NSSE on an every other year cycle. Interestingly enough in the administration of last cycle of NSSE—2006, all five of the national benchmark scores for SC were higher than the corresponding benchmark. Savannah College of Art and Design Savannah College of Art and Design (SCAD) is a private master’s level institution focused on art, design and architecture, with campuses in Savannah and Atlanta, Georgia. The 2006 enrollment comprised over 8,900 students from 108 countries.

Use of assessment findings for curricular reform Method Founded in 1978, SCAD has seen extremely robust growth over its 29 years as a college. The rapid growth of the college has meant that an intuitive managerial and decision style that was functional at a smaller enrollment size has gradually given way to more quantitatively oriented decision making. A recent over-funding of institutional scholarships has forced the college to re-evaluate the basis upon which students are awarded institutional grants and scholarships, to ensure that those most deserving and of need in fact benefit. Utilizing student demographic data, high school performance information, data on financial aid awards and demonstrated need, and SCAD student performance data (defined here as student persistence, progression, grade point average, and graduation), researchers in the Office of Institutional Research have sought to identify whether or not any effect exists for the awarding of institutional aid using the 2000 incoming cohort of first-time full-time freshmen. At present, logistic regression models have been used to study student persistence/ retention and graduation. Results While the study of the effect of institutional scholarship awards is ongoing and final results have not yet been obtained, initial results point out a distinctive difference in impact on student graduation by level of institutional scholarship offered. Small institutional scholarships have not been found to be effective. Surprisingly, institutional scholarships that exceed US$40,000 over the total period of attendance have also not been found to have a substantial positive impact on graduation. Moderate-sized levels of institutional scholarship, on the other hand, have been found to increase the log odds of graduation by factors ranging from five times greater than no institutional aid to 23 times greater than the absence of institutional aid. Preliminary results are now being validated against the 2001 incoming student cohorts. The institutional impact has been immediate: templates have already been developed in anticipation of the study’s findings to guide admissions counselors as to the maximum institutional award from which a student is likely to benefit, based on factors identified as statistically significant. This template should therefore maximize the efficiency to the institution of the scholarships and awards granted, and should enable the institution to compete more effectively for high-quality students in its niche marketplace. The use of institutional research in large doctoral research institution The second case study will compare the use of data and information developed through institutional research for decision making from the standpoint of large public research institutions. North Carolina State University NC State is a large technology-intensive doctoral research university with over 31,000 total students enrolled in 2006, and $207 million in sponsored research funds during fiscal year 2006. The institution is located in Raleigh, North Carolina, and is part of the University of North Carolina System. Admissions index Method Using the performance of the immediately previous freshman cohort as a basis for analysis, institutional researchers at NC State have developed least-squares regression models to predict the academic performance of the incoming freshman student body. Set in the context of multiple-year enrollment management plans, the context for this study was a thorough knowledge of student retention, progression, and graduation by program. Factors such as high school grade point average, standardized test scores, years of enrollment in foreign language courses, and scores on an extracurricular activity index were used develop an overall university and separate college-level models of predicted first-year student academic performance. These models were recalibrated on an annual basis. Results Predictive modeling of student first-year performance resulted in statistical models that explained an average R2 of 0.25 at the university level over the years 1993–1997. The existence and quality of these models led to the adoption of enrollment quality targets, jointly decided in enrollment planning meetings between institutional research, admissions, and academic leadership in the colleges. Informed by an institution-wide 10-year enrollment growth forecast, and armed with an awareness of their current throughput and output capabilities, college administrators were able to predict the number of incoming students at a given quality and likelihood of success (graduation) they were able to accept into their programs, and plan resource allocations both efficiently and effectively to accommodate the incoming student body.

Project 25 Method Project 25, a mixed-methods project undertaken by researchers in University Planning and Analysis, was initiated in late 1997 to develop an understanding of the impact on students and faculty of providing online sections of 25 university courses at the graduate and undergraduate level to currently enrolled full-time students (see http://legacy.ncsu.edu/info/f97\_assessment.html). The study included student demographic and performance (grade point average) data; survey data from students enrolled in parallel sections of online and on-ground courses at NC State, student focus groups, and structured interviews of faculty who developed and taught the parallel online/on-ground sections of the courses studied led to the　provision of dedicated resources for faculty development of online courses. Results Results confirmed the lack of a statistically significant difference in student performance between the parallel online and on-ground course sections studied. The same lack of significant differences held by gender. Students participating in focus groups noted that heightened interactivity between professor and student in the online environment made a positive difference. The findings of this study formed the empirical and policy groundwork for the expansion of online offerings at the institution. Faculty who developed the online course sections noted the huge time commitment necessary to develop parallel online sections of extant courses, and recommended that the university develop a resource base for faculty development of online courses. In response, the university initiated a support presence that has evolved into the current Distance Education and Learning Technology Applications center (see http://lts.ncsu.edu/) that provides a rich and deep resource for faculty to teach with technology, itself a development that enables the university to leverage institutional resources to increase its outreach and enrollment of online students in the United States and elsewhere.

Georgia Institute of Technology

Georgia Tech is a preeminent engineering and technology focused doctoral research institution in Atlanta, Georgia that enrolled 17,900 total students in 2006 and had $347 million in sponsored research expenditures for fiscal year 2006. Georgia Tech is a constituent institution of the University System of Georgia. President’s Undergraduate Research Awards (PURA) Program Method In spring 2000 and spring 2001, Georgia Tech undergraduate first-year and senior-year students participated in early administrations of the NSSE, jointly administered through Institutional Research and the Office of Assessment. The NSSE is largely concerned with student self-reports of activities that have been frequently mentioned in the literature base as having a positive effect on student outcomes. One particular item in the survey was of high interest to the president of Georgia Tech: the proportion of undergraduates who reported involvement in an undergraduate research project. Results The collection of NSSE survey data from senioryear students during spring 2000 and spring 2001 led to the finding that far fewer students were undertaking undergraduate research projects than previously thought. Subsequent follow-up research confirmed the negative role of factors such as a lack of faculty incentives to mentor undergraduate students in research projects and a de facto penalty on the faculty in terms of lost professional research time for involving themselves in directing undergraduate research projects. The immediate impact of this finding was the creation of the President’s Undergraduate Research Awards (PURA) program, funded annually at US$250,000. This program provides competitive grants for students and faculty to undertake undergraduate research projects. The longer-term impact of the NSSE survey findings has been the creation of the Undergraduate Research Opportunities Program (UROP) at Georgia Tech as part of the institution’s reaffirmation of regional accreditation in 2005 (see http://www.undergradresearch.gatech.edu/). Taken together, these programs have steadily increased the proportion of undergraduates reporting engagement in research projects. According to the UROP newsletter, in 2006 alone 1,790 students enrolled in one of the more than 100 undergraduate research courses at Georgia Tech, a figure that represents an increase of over 34% from the previous year (see http://undergradresearch.gatech.edu/documents/ UROP-news-August%202007-ver2.pdf).

The use of institutional research in a state system

The third case study will compare the use of data and information developed through institutional research for decision making from the standpoint of a state university system. California State University

The California State University (CSU) is a leader in highquality, accessible, student-focused higher education. With 23 campuses, over 430,000 students and 46,000 faculty and staff, the CSU is the largest, the most diverse, and one of the most affordable university systems in the country. The CSU derives its role in California higher education from the California Master Plan for Higher Education. The Master Plan envisioned a three-part postsecondary system that would provide Californians with world-class research, free (or nearly free) access to higher education for all, and the full complement of skilled workers, professionals, researchers, managers, executives, and other leaders to make California the best that it could be for all Californians. The University of California (UC) was created to serve as the primary public research university for California and has been asked to serve the top eighth of high school graduates and to provide master’s and doctoral programs in all disciplines and professional areas. The California Community Colleges (CCC) is the public postsecondary entity with the duty to provide all Californians with the opportunity to learn, to transfer to a baccalaureate granting institution, and to gain vocational skills through certificate and Associate of Arts programs. The CSU is the middle partner in the Master Plan.

The CSU has been asked to serve the top third of high school graduates and together with the CCC essentially to comprise the transfer function in California, providing instruction and degrees to over three-quarters of CCC transfers to baccalaureate programs in the State. In addition, the CSU offers master’s degrees and joint doctoral programs. Since its creation in the early 1960s, the CSU has served as the economic backbone of the State, providing not only teachers, but professionals in accounting, marketing, small business, agriculture, criminal justice, public administration, and tourism. As the world has flattened and applied learning has become increasingly important, the CSU increasingly is playing strategic roles in applied research and has just launched its first independent applied doctoral program in educational administration and leadership to address one of California’s most pressing needs, the need for educational leaders in public K-14 education. California has been the most diverse state in America for decades, and the CSU has needed to embrace the educational challenges of enabling diverse, old and new, Americans to seek and reach their dreams. The public K-14 education has been the foundation on which California has depended, and it requires strengthening. Expectations of the CSU system are high—from the desires of the government that still provides the vast majority of capital and operational funds, to business and industry who depend on CSU graduates for skilled, yet nimble and adaptive, professionalism, to the families and children of California who want the very best, cost-effective, opportunities for participating in life in California and, indeed, in the world. The role of the CSU system is to stay abreast of the evolving challenges and expectations, to undertake initiatives to address these challenges and expectations, to get the requisite resources, to allocate the resources to CSU campuses to address the challenges, and to account to those who fund, depend upon, and need the CSU. Perhaps most importantly, the system recognizes that it is only at the campus where faculty, students, administrators, and staff actually do the fundamental business of the university— teaching and learning, connecting with local communities and regional businesses and industries, and engaging and enhancing students’ lives. The cases discussed above focus on the ‘‘real’’ business of the university and improving it. The cases below deal more with the role of institutional research in the strategic initiatives to address evolving trends in higher education and the use of institutional research. The eligibility of California public high school graduates for the CSU and the UC Method The California Master Plan for Higher Education requires UC and CSU to establish clear admissions requirements that guarantee access to California high school graduates that satisfy these requirements. Students that have satisfied these requirements are deemed ‘‘eligible’’ to attend UC or CSU and are admitted if they satisfy the relevant application requirements. The Master Plan specifies that UC’s requirements shall make eligible the top 12.5% of public high school graduates; CSU is required to make eligible the top 33.3% of these graduates. To the extent that UC or CSU requirements do not make appropriate percentages of graduates eligible, it is expected that the systems will amend their requirements. This is one of the most high-stakes ways in which institutional research impacts decisions in higher education. Every California high school graduate who is CSUeligible must be afforded access as a first-time freshman to at least one CSU campus. This is the promise of the State. As part of its accountability to the State and its students and families, institutional research assesses the fit between the promise and its requirements. When necessary, institutional research, then, provides data-driven options by which the system’s consultative governance structure makes decisions about adjustments to eligibility coursework, grades, and test scores. Eligibility studies have been conducted about every 4– 6 years since the Master Plan was enacted in 1960. The most recent study was conducted for the class of 2003. The next study is planned for the class of 2007. The research plan for the last study and the next has been to sample California public high schools and evaluate the transcripts of all seniors at each of the selected schools. The number of high school graduates serves as the denominator for all eligibility rates. Eligibility rates are generated for by gender and for four ethnic categories (i.e., White, Black, Latino, Asians, and others). Official eligibility rates are based on students that attended comprehensive, alternative, and continuation high schools. Eligibility rates for students that attend comprehensive schools also will be generated for comparative purposes. The research design and methodology for the last and next eligibility studies differ markedly from previous eligibility studies. In previous studies, a few transcripts from every public high school in the State were collected and evaluated—because it was felt that high school students, their parents, school counselors and administrators, and school boards would not accept the results of eligibility studies if none of their graduates’ transcripts were included. In the 2000s, increasingly students apply online for admission to college. Self-reported courses, grades, and test scores provided online are used effectively by the CSU and the UC to provide provisional admission offers for many anxious prospective freshmen. These offers are followed up by review of final official transcripts and test scores. In addition, statewide K-12 tracking systems have been under development and in the future offer the promise of electronic transcript review from every high school in America. Student use of online applications and increasing use of and comfort with electronic transcript storage by public high schools essentially permitted a transition with the 2003 study to a less manual approach. By the 2003 eligibility study (CPEC 2003), 400 of all 1,005 public high schools that maintained electronic transcripts maintained in a SASIxp format (and comprised 40% of California high school graduates). From the 400 schools, nearly 16,000 transcripts were collected from 48 schools. A sampling and statistics consultant was hired to guide initial planning, and analyses were performed first to confirm that the 400 public high schools were not unlike the full population of public high schools in terms of academic performance, academic achievement, and actual university enrollment rates, and, then, to ensure that the actual 48 high schools used in the study met these requirements as well as others. Vangent (formerly Pearson), an outside vendor, had and have four major data processing and management tasks to perform. First, it is responsible for developing software products that will read electronic transcripts and provide functionality for determining whether transcript meets either CSU or UC eligibility requirements. Second, it is responsible for collecting all the sampled transcripts and matching SAT/ACT scores to each student’s record. Third, it is responsible for creating comprehensive course listings for each school and for creating sets of college preparatory course A-G status codes (and ‘‘other’’) for each course listing associated with each sampled school. Fourth, it is responsible for making the initial determination on whether a transcript represents a CSU- or UC-eligible student, and it is responsible for incorporating feedback from CSU and UC transcript analysts to make the final determination on whether a transcript represents a CSU- or UC- eligible student. Vangent is also responsible for functioning as a liaison for each high school and providing technical support to CSU and UC staff throughout the study. The CSU and UC transcript analysts are responsible for verifying that the course status codes generated by Vangent. This means that the analysts must evaluate the comprehensive course listings of each sampled school. The main problem frequently addressed by the analysts is the ‘‘true’’ A-G status of transfer courses (i.e., courses not administered at the same school where students earned their diplomas). As should be clear from the foregoing, the obvious reason why all 400 electronically enabled high schools were not included in the study was the need to classify and confirm every unique high school course into a course status code. For the UC in 2003, many ‘‘pathways’’ to eligibility for admission exist. For the CSU, all college preparatory courses must be taken with at least a passing grade; in 2003 an overall GPA of 3.0 plus completion of the courses afforded a student with CSU eligibility. If such a student did not have an overall GPA of at least 3.0 an index involving the GPA and SAT/ACT score may be applicable. While transcript analysts are essential to the study, eligibility studies themselves are overseen by institutional researchers who review and make the research study design and sampling decisions offered from among the consultant’s options. Results From 2003 study, the CSU learned that 31.3% of the comprehensive public high school graduating seniors (±3.4%) were eligible for the CSU. Males were at 26% eligibility, females at 36.2%; African Americans at 20.7%, Asian Americans at 49.2%, Latino Americans at 17.6%, and Anglo Americans at 37.2%. Results from adjustments that included estimates regarding alternative and continuation school graduates suggested an overall 28.8% eligibility rate for the CSU. The CSU, unlike the UC, focused its policy deliberations on the results of the actual transcript study—that is, on the results of comprehensive high school graduates. The study was framed to focus on public comprehensive high schools and it is toward public and private comprehensive high schools that system and campus efforts are made to ensure improving knowledge about college preparatory course requirements, reviewing that high school are offering college preparatory courses, and improving the quality of secondary school pedagogy (especially in mathematics). Because the 2003 CSU eligibility rate was within the confidence interval of 33.3%, no adjustments were made to the eligibility criteria for firsttime freshman admission. However, you will recall that the 95% confidence band for the eligibility rates was ±3.4% for the CSU in the 2003. The 1996 study had a 95% confidence band of ±0.8%. Criticisms were raised by many, but especially with regard to UC eligibility estimates for underrepresented minorities where the African American measure of central tendency 6.2% was smaller than its confidence bandwidth (6.8 percentage points). Obviously, the systems’ 2003 research informed the decisions about the 2007 study. The 2007 will quadruple the number of high schools in the study sample, and quadruples the cost to over $1 million which will be allocated to Vangent and transcript analysts. The operation of the 2003 eligibility study brought further dialog between the CSU and the UC system academic senators and admissions personnel to align college preparatory course requirements within categories, like visual and performing arts and the sciences—activities that build upon previous adjustments. The faculty academic senators and admissions personnel recognized the importance in aligning basic college preparatory requirements to facilitate student, family, and college adviser understanding and actions and acted on it, as they had in the past. The 1996 eligibility study results revealed that a small proportion of very bright high school graduates were eligible for the UC, but not for the CSU, because the CSU had the additional course preparatory requirement of a course in visual and performing arts that was not being met; college counselors were advising college seniors to take an extra mathematics or science course to improve their chances of gaining admission to one of the more competitive UC campuses. The 1996 results encouraged the CSU academic senators and administrators to confer and meet with their UC counterparts finally to require the same college preparatory course areas for UC and CSU eligibility. Finally, it is anticipated that the CSU eligibility rate may drop below the confidence band. Should this occur, institutional researchers will model various choices for the CSU admissions advisory council to review. This advisory council is headed by a CSU president and involves a CSU provost, a CSU student affairs vice president, CSU student association representatives, and a number of CSU academic faculty senators. In theory, the admissions advisory council will have a broad range of choice: adjusting the eligibility index (a linear regression of high school GPA plus optional tests [SAT/ACT]) to permit students with lower than a 3.0 GPA to gain admission by scoring at lower levels than are currently required CSU eligibility, lowering the ‘‘automatic’’ 3.0 GPA, or deleting a college preparatory course requirement. In addition, some have suggested (and the 2007 eligibility study design has been amended to address it) that the implementation of the California High School Exit Examination (CAHSEE) has altered the traditional pool of UC/CSU eligible students; it is anticipated that some high school seniors who have taken all the college preparatory courses and are otherwise eligible for the UC or the CSU did not pass the CAHSEE and, thus, did not receive a high school diploma and are therefore not eligible for admission. The UC and the CSU have supported the State Superintendent of Schools, the State Board of Education, the District Superintendents, and District Boards in requiring a high school diploma of all entering first-time freshmen. The council’s recommendation for adjustments would go to the system Chancellor for his deliberation and action. To the extent that the CSU is underachieving its Master Plan assignment, no doubt exists that adjustments will be made, and there is no doubt that CSU and their partners in the media, community organizations, and public schools will translate the adjustments into English and heritage languages to get the word out clearly and quickly. It also is worth noting expanding the eligibility pool to students who are somewhat less prepared academically will increase the need of CSU first-time freshmen for remediation—beyond the half who now requires extra assistance in mathematics and English. NASULGC/AASCU voluntary system for accountability Method Participant observation is the method for this case study, although it should be clear that without the 50 years of institutional research, there could be no initiative among public higher education institutions broadly to account. Under the Bush administration, the No Child Left Behind initiative created a framework for monitoring state K-12 education and a similar effort was initiated in combination with the need to reauthorize the Higher Education Act. There have been calls for a unit-record tracking system of all students in postsecondary education in order to monitor the extent to which students are progressing to degree and costs of postsecondary education are reasonable. There have been calls for national tests of student learning in higher education, much as elementary and secondary school students are tested and results posted. The Association for Institutional Research (AIR) itself and individual institutional researchers representing their universities have participated in several years of discussion and debate about the broad goals and the devilish details of data collection, specification, and reporting. While Congress and the Department of Education have not yet created circumstances to finalize and implement major new notions of higher educational accountability, it seems clear that even with a change in presidential administration, the federal government will continue to have concerns about the costs of college, student indebtedness, degree completion, and student learning. A Test of Leadership: Charting the Future of U.S. Higher Education (Spellings Commission 2006) summarizes the call for increased access to higher education; a restructured financial aid system; consumer-friendly information, accountability, and transparency (including the measurement and reporting on value-added student learning outcomes); innovations in pedagogy and programs to meet the changing needs of a global knowledge economy and for lifelong learning. The CSU’s Early Assessment Program (EAP) was identified as one of the best national models of how higher education and K-12 officials can collaborate to help students prepare for college (Spellings Commission 2006, p. 16); the EAP involves the confirmation of alignment between K-12 standards and CSU entry-level standards in mathematics and English, the use of the regular 11th grade California Standards Tests (CSTs) in mathematics and English (with short augmentations) to provide end-of-year high school juniors with a signal before the start of the 12th grade about their readiness for college-level mathematics and English, EAP professional development for high school English and mathematics teachers, the revision of pre-service preparation for mathematics and English teachers, online CSU Math and English Success websites for students as early as middle-school to chart curricular pathways to success, to get more diagnostic information, and to engage in online mathematics and English instruction; system institutional research staffs the EAP tests that are built on the foundation of the CSTs and is responsible for creating the framework for ongoing evaluation of precollegiate interventions. Special mention is made in the report of the Collegiate Learning Assessment (CLA), the National Survey of Student Engagement (NSSE), the Community College Survey of Student Engagement (CCSSE), and the National Forum on College-Level Learning with regard to examples of student learning assessments (Spellings Commission 2006, 23). The boards of the National Association of State University and Land-Grant Colleges (NASULGC) and the American Association of State Colleges and Universities (AASCU) decided that they would attempt to create consensus among member institutions to address many of the recommendations of the Spellings report, especially the need for consumer-friendly information about higher education. While the acronym DRIP—data rich, information poor—has its roots in business and information technology, DRIP is exactly what characterizes the concern for consumer- friendly information. The 50 years of Institutional Research helped to support the development of a complex, large federal data collection system on postsecondary education collection—the Integrated Postsecondary Education Data System (IPEDS). Every postsecondary institution receiving federal financial aid annually must submit summary aggregate information (frequently disaggregated by race/ethnicity and by gender) regarding student admissions, student enrollments, student financial aid, costs to students, faculty and staff, program enrollments and completions, retention, and graduation, as well as excerpts from audited financial reports. IPEDS has a number of tools for extracting single or multiple years of data on selected institutions and across a myriad of variables. The data are rich (although some bemoan their limitations), and there have been federally financed attempts to present more consumer- friendly information. Groups, like the Ed Trust have develop College Results Online (http://www.college results.org/) from IPEDS data to help people learn more about the student graduation rates at various colleges and universities. The National Center for Public Policy and Higher Education has mined IPEDS data and other public sources of information to develop its annual Measuring Up: The National Report Card on Higher Education (http:// measuringup.highereducation.org/nationalpicture/). Measuring Up provides simple graphics on national and state-by-state trends in preparation for college, college participation, college affordability, college completion, college benefits, and learning. NASULGC and AASCU decided to see if they could join together to provide ways of helping their member institutions provide consumer-friendly information that would help parents and students understand more about their colleges and universities’ admissions, affordability, programs, campus life, life-after-college, student engagement, and student learning. The NASULGC board of 28 includes four from the CSU, including the CSU System Chancellor. The AASCU board of 16 includes three from the CSU. The CSU System Chancellor and CSU Fresno President serve on the Presidents’ Advisory Council to the NASULGC/AASCU Voluntary System of Accountability (VSA). To inform the VSA, two CSU presidents agreed to chair technical working groups on learning outcomes and student growth. A CSU provost served on the student engagement task force; a CSU vice president of student affairs served on the working group on student growth and another on student and family information task force; a CSU vice provost served on the learning outcomes work group. A system design and information task force was added to mix, and the CSU system assistant vice chancellor for academic research who heads of the institutional research area was invited to participate along with two additional institutional researchers. In short, the CSU system committed executive and operational leadership to the VSA to shape the initiative. To complement the NASULGC/AASCU VSA, the CSU system developed its own CSU Presidents’ Council on Accountability (CSU PCA)with nine CSU presidents, including the three on VSA. The CSU PCA, along with the CSU System Chancellor, committed that every CSU would experiment with the Collegiate Learning Assessment (CLA) with basic funding provided from the system office. The CSU system assistant vice chancellors for academic programs and academic research convened a north and a south workshop for the 23 CSU campus administrators who were tasked to undertake the experiment; the workshops included detailed presentations from the creators of the CLA, as well as the chief implementation officer. CSU campus administrators who already had experimented with the CLA shared their observations and tips for implementation and use of results. Results The task forces for the VSA have reported, and the VSA President’s Advisory Council has recommended implementation of the VSA to the boards of NASULGC and AASCU which will be meeting in November 2007. Consultation in each of the work groups and task forces was broad and responsive. Three student learning outcome measures have been approved for use by those who agree to join the VSA initiative. The measure mentioned in the Spellings report, the Council for the Advancement of Education’s Collegiate Learning Assessment (CLA—http://www.cae.org/cla/) is one of the approved student learning outcome measures. Another is the ACT’s Collegiate Assessment of Academic Proficiency (CAPP—http://www.act.org/caap/). The third is the Educational Testing Service’s (ETS’) Measure of Academic Proficiency and Progress (MAPP—http://www. ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b 13bc3921509/?vgnextoid=ff3aaf5e44df4010VgnVCM100 00022f95190RCRD&vgnextchannel=f98546f1674f4010 VgnVCM10000022f95190RCRD). Four surveys of student experiences and perceptions were approved by use for those who agree to join the VSA initiative. The survey mentioned in the Spellings reports, the Indiana University Center for Postsecondary Research’s National Student Survey of Engagement (http:// nsse.iub.edu) is one of the surveys. The second measure is the College Student Experiences (CSEQ), a research program originated by C. Robert Pace, UCLA Professor Emeritus and Humbold State University legend and housed at the Indiana University Center for Postsecondary Research. The third is the UCLA Higher Education Research Institute’s (HERI’s) College Senior Survey (CSS) which is part of HERI’s Cooperative Institutional Research Program (CIRP)—http://www.gseis.ucla.edu/heri/css.html. The fourth is the University of California’s Undergraduate Experience Survey (UCUES), housed at Office of Student Research at UC Berkeley (https://osr2.berkeley.edu/Public/ surveys/ucues/menu.html). Most impressively, the task force on student and family information which included one system institutional researcher and several campus institutional researchers took the lead in developing the principles that the VSA initiative ought to minimize institutional burden and be implementable by a one-person institutional research office. To this end, an EXCEL workbook template based on using the Common Data Set (CDS—http://www. commondataset.org/, listserve hosted by the College Board) loads automatically to populate most of a report that can be posted or printed was developed and is being tested by a number of campuses (including several in the CSU). One major enhancement to the VSA student and family portion of the template is the definition of Undergraduate Success and Progress. First, unlike the IPEDS collection, transfer student success and progress also are displayed. More importantly, a partnership with the National Student Clearinghouse, one of the nation’s most trusted sources for student degree and enrollment verification (http://www. studentclearinghouse.org/) is entering into an agreement with the VSA to provide each VSA institution with retention and degree completion statistics for their entering cohorts of freshmen and transfers to institution’s beyond the original, consistent with the work of Clifford Adelman’s work that showed that students make their way to degree through a variety and number of institutions (Adelman 1999). VSA wants to drive home to parents and students that it is not unusual for a student to being college at one place and move to another as they progress to degree. Prospective students and parents ought to know that circumstances frequently change, and making progress and succeeding really are what is important. One’s original institution is the place of entry. VSA also committed to having a cost-calculator URL attached to the online posting and to facilitate institutions, one is being developed by one of the participating institutions. The purpose of the cost calculator is to let parents and students know that that a college sticker price is just the beginning, and, in fact, college probably is within their budget because of federal, state, and institutional financial aid. The plans of graduating seniors also are a required addition beyond the CDS loading. Additional spaces and pages of the VSA’s College Portrait present the required student experiences and perceptions outcomes from one of the approved surveys and the required student learning outcomes indicators from one of the approved measures. There is a timeline for required reporting of survey outcomes and learning outcomes—most within 2–4 years of signing onto the VSA. The VSA Presidents’ Council has approved moving a final recommendation to implement VSA forward at the NASULGC and AASCU meetings in November 2007, and, at least, two CSU campuses will be part of the national rollout. While the NASULGC/AASCU VSA has yet to be implemented, there is preliminary word that the Department of Education and others in Washington, DC are pleased with the efforts. Excerpted from a NASULGC notice on September 28, 2007 (http://www.nasulgc.org/ NetCommunity/Page.aspx?pid=763&srcid=183) is the following: ‘‘The Association of American Colleges and Universities (AAC&U) was awarded a grant from the U.S. Department of Education’s Fund for the Improvement of Postsecondary Education (FIPSE) to support an initiative on student learning assessment. ….Rising to the Challenge: Meaningful Assessment of Student Learning establishes a consortium among AAC&U, the American Association of State Colleges and Universities (AASCU), and the National Association of State Universities and Land-Grant Colleges (NASULGC) to collectively build campus leadership and capacity to implement meaningful student learning assessment approaches and use assessment results to improve levels of student achievement….The effort will be led by AAC&U Vice President for Quality, Curriculum and Assessment Terrel Rhodes, with NASULGC Vice President for Academic Affairs David Shulenburger, and AASCU Director of Special Projects and Development John Hammang rounding out the leadership team. AAC&U, AASCU, and NASULGC collectively comprise more than 1,360 distinct member institutions (public/private, large/small, 2- and 4-year, liberal arts/comprehensive/research). The project responds to the need to examine the multiple purposes of learning assessment and to test the validity, comparability, and appropriate uses of a variety of assessment approaches. It builds on the existing learning outcomes framework provided by AAC&U’s Liberal Education and America’s Promise (LEAP) initiative, the AASCU/NASULGC Voluntary System of Accountability (VSA) project, and AASCU’s American Democracy Project. The joint initiative will support three main strands of work: NASULGC will coordinate efforts by educational researchers and leaders from the Educational Testing Service (ETS), the Council for Aid to Education (CAE), and the American College Testing Program, Inc. (ACT) to examine the extent to which disparate measurement tools recommended as part of the VSA can be used interchangeably, whether these tools are measuring similar or dissimilar outcomes or levels of achievement, and the role test format (e.g. multiple choice vs. open-ended/constructed response measures) plays in the correlation among measures. AAC&U will lead an effort to develop an e-portfolio framework for assessing a wider array of learning outcomes than those measured by these other tests. This part of the project will foreground practices that base assessments on authentic examples of student work collected over time in an e-portfolio. Titled VALUE, this research and development effort will collect and synthesize best practices in facultydeveloped rubrics to highlight commonalities of outcomes and expectation of achievement levels across institutions. AAC&U also will develop models and templates through which e-portfolios can be used to demonstrate, share, and assess student accomplishment of advanced and integrative learning outcomes. AAC&U’s VALUE project was first launched earlier this year with support from the State Farm Companies Foundation. AASCU will lead a third part of the initiative to develop a validated survey instrument to measure changes in student growth especially related to the development of competence in skills effective in the workplace and those related to civic engagement. Rising to the Challenge brings together three leading national higher education associations around a complementary set of shared commitments to generate more information, but also to raise the level of student achievement. It seeks to help campuses use effective tools that provide meaningful information about how well students are learning essential learning outcomes important for work and life in a complex and rapidly changing world.’’ CSU faculty, administrators, and institutional researchers will be involved in, at least, two phrases of the Rising to the Challenge. AAC&U plans to involve CSU general education assessment leader and specialist in e-portfolios, as well as one or two disciplinary leaders in e-portfolios. AASCU has invited three CSU faculty and deans in non- English language proficiency, civic engagement, and the workplace to join its panel of about ten content experts. Two CSU institutional researchers will participate with five other colleagues nationally in validating the growth survey with a coordinating hand from the CSU system office. Inside Higher Education (2007) in its reporting on the $2.4 million FIPSE award said, ‘‘Officials of the Education Department.…herald the development as a breakthrough…. The embrace of the project by the three groups….shows that colleges are unafraid to assess their performance, but want to ensure that the measures used are appropriate, intelligently crafted, and fully represent the many kinds of skills that students need.’’

Discussion

The intent of the authors in compiling these case studies is to compare and contrast the roles of Institutional Research in informing decision making and governance in higher education. Clearly these cases illustrate the effective use of Institutional Research across the different sectors of higher education: whether it is the collaboration of institutional research with various other offices and constituents illustrated in the case studies from the private sector; the effective use of analytical tools to address emerging issues illustrated in the case studies from the large doctoral institutions; or the effective use of public communication and accountability illustrated in the system level case study. Clearly, Institutional Research stands at the nexus of collecting, analyzing, and providing consistent, reliable information, but more than that Institutional Research is a vital component in effective governance of higher education. Our cases illustrate that Institutional Research supports the assessment of student learning, informs institutional management, and supports strategic planning. The rise of accountability in higher education is not just a passing fad. Institutional Research offices are most often responsible for the data that is provided at the institutional, system, and national level. Public demand for transparency and accountability has increased the demand for Institutional Research. Our cases document the quality of reports and information from offices of Institutional Research that are filling this public demand. In addition our cases illustrate the need for more than just the reporting of facts and descriptive statistics to understand the complex phenomena we know as higher learning. Most often facts and descriptive statistics raise as many questions as they answer. As a result, Institutional Research is increasingly called upon to use multiple methods of assessment, a variety of research designs, and advanced analytics to understand and document effectiveness of our institutions. Institutions of higher education are currently competing for students in a global marketplace (Friedman 2005). As a result of this competition, the need to provide quality programs has brought to the forefront the role of assessment. Our cases illustrate how assessment data is used in both a formative and summative fashion. Within the case study examples from SC and SCAD, the institution used the results of their research to inform practices and changes in their academic programs and the quality of the experiences of their students. Yet, our cases also illustrate how institutional research brings data to the public. The cases from the CSU document how Institutional Research serves to keep California’s promise of access to its students and families, while also leading a national effort to inform the public about our institutions of higher education. In informing the public about our institutions, these Voluntary Systems of Accountability become a marketing tool, as they are used by future students to explore college opportunities. Clearly, all institutions of higher education are in a competitive position in the recruitment and retention of students. So as these data are brought forth for public consumption, it becomes a summative assessment of the institution as a whole that is often used by future students as they make their college choices. The above trends exemplify some of the common grounds upon which institutional research informs governance in higher education across the different sectors. While we share common grounds, differences do exist across the sectors. At the system level the main focus is on accountability, while at both public and private level, Institutional Research tends take more of an institutional view for management purposes. Public institutions by definition have more transparency due to revenue streams, while private institutions by definition have less transparency. In private institutions accountability is more based upon institutional stakeholders (e.g., students, parents, and faculty). Other similarities shared across sector include the relationships between the local municipalities and our institutions. In addition, all institutions are tied together across curricular disciplines. In fact, across public and private sectors, institutions that share similar program mixes are more similar to one another than the institutions within the same sector with very different programmatic mixes. Also, institutions are brought together across accrediting agencies for curricular issues. Thus, effective Institutional Research supports the assessment of student learning, informs institutional management, and supports strategic planning. Without such support effective governance within higher education is not feasible. While sector differences do exist in the role of institutional research in informing decision making and governance in higher education, more similarities than differences exist.

Over the last 20 years, the UK higher education sector has become increasingly competitive. The entry of new institutions into the university sector in 1992, the growing internationalisation of higher education and the decrease in central government funding (in terms of unit of resource) has forced institutions to devise and establish strategies for maximising their sources of income. This period has also seen the increasing influence of league tables, both national and international, and the association of reputation with ranking. Within this environment, research has progressively been seen as an activity that, if well managed, can enhance an institution‟s reputation and international standing and also generate potentially high economic returns. It is within this context that research activity can be described as an institution‟s research capital. (As the next paragraph makes clear, the use of the term 'research capital' is more specific to academic activities than the manufacturing/service activities normally embraced by the traditional, economic-based use of the term. See Griliches, 1986, or Mansfield, 1980, for the latter definition.) The core of what constitutes the institution‟s research capital is well defined and understood by the academic community, higher education senior management and funding agencies. Research activities such as the production of publications/outputs, the ability to attract external funding for research (whether from public or private organisations) and the training of future researchers through post-graduate research programmes are all at the core of an institution‟s research capital. Moreover, the boundaries and definitions of research capital have been further consolidated over the last 20 years, through a number of national research assessment exercises and the programmes of research funding agencies. However, a change in the political landscape has seen a new emphasis placed on the measurement of the socio-economic impact of publicly funded research. This emerging theme was a corner-stone of the government‟s 10-year (2004-2014) plan for science and innovation (HM Treasury, Department for Education and Skills, Department of Trade and Industry, 2004). In October 2006 a joint Office of Science and Innovation (responsible for co-ordination of the Government's science and technology policies) and Treasury Steering group was formed and subsequently the recommendations from this group developed into the former Department of Innovation, Universities & Skills – DIUS (now the Department for Business, Innovation & Skills – BIS) investment framework (DIUS, 2008). This sets out a reporting structure, the Economic Impact Reporting Framework (EIRF), for the Research Councils UK (RCUK). The growing importance of socio-economic impact can also be seen in the new proposals from the Higher Education Funding Council for England (HEFCE), which also reports to BIS, to introduce the assessment of socio-economic impact as part of their Research Excellence Framework, the replacement for the well established Research Assessment Exercise, that was last deployed in 2008 (HEFCE, 2009). The challenges of capturing and assessing socio-economic impact are not new either in the UK or internationally. A recent report by RAND Europe (Grant, Brutsher, Kirk, Butler, & Wooding, 2010) identifies 14 different frameworks to assess impact from Australia, the United States, the Netherlands, Sweden, Canada, Japan and the UK (the most notable of which are: the Australian Research Quality and Accessibility Framework, the US Program Assessment Rating Tool and the Dutch Evaluating Research in Context). The challenge for institutions with a broad disciplinary base, given the breadth of definitions of socio-economic impact, is how to define, identify and capture impact systematically for different disciplines and different types of research. Even more important is the challenge to devise strategies for actively supporting and facilitating the pathway from research to impact. The current body of knowledge on the impact of academic research gives an insight into how management strategies can be developed to integrate socio-economic impact into the institutional research capital.

Current understanding of the impact of research

Many funding agencies and charities have embarked in impact studies including: The British Academy in January 2004, the Medical Research Council in collaboration with the Wellcome Trust and the Academy of Medical Sciences in 2008 (Health Economics Research Group, Office of Health Economics, RAND Europe, 2008), and the Arthritis Research Campaign in 2009 (Wooding, Nason, Starkey, Hanney, & Grant, 2009). Most notably, the Economic & Social Research Council (ESRC), one of the seven councils that make up the Research Councils UK, commissioned in 2006 a series of impact case studies aimed at identifying the impact of the research it had funded (ESRC, 2009). Impact studies use a number of methods and tools to identify and evaluate impact. Several of these methods, such as the Payback Framework developed at Brunel University (Buxton & Hanney, 1996; (Buxton, Hanney, & Jones, 2004), define impact on the non-academic community as one of the many dimensions of research impact. In these evaluative frameworks, knowledge generation (academic impact), capacity building (training and career development), and socio-economic impact (impact on policy, practice and the wider community) are all different expressions of the value of research. This view acknowledges that research might generate different types of impacts, at different times and with different intensities, and asserts that by capturing all these dimensions of impact we will have a true representation of the value of the research. Conceptually this view has been reinforced by the DIUS Economic Impact Reporting Framework which is structured around a number of categories (dimensions) that define economic impact (DIUS, 2008). These categories are:

Overall economic impact: Increased productivity

Improved welfare

Innovation outcomes and outputs: Technological innovation

Wider innovation

Knowledge generation: Human capital

Stock of publicly available knowledge

Investment in the research base and innovation: Expenditure on R&D

Expenditure on innovation.

Table 1 - Economic Impact Reporting Framework Categories

It is interesting to note that the Economic Impact Reporting Framework includes all aspects of academic research as activities that result in economic impact, including what we might define as the research capital mainly aimed at the academic community. This interpretation of economic impact re-affirms the concept that socio-economic impact is not a distinct outcome, sometimes perceived as in conflict with the „pure‟ creation of knowledge, but can be seen as just another dimension of academic research. From an institutional perspective, therefore, socio-economic impact should be seen as an addition to the research capital that can be captured and harnessed as one of the many outcomes of academic research activity.

Institutional challenges

While the majority of the impact studies conducted by funding agencies might vary in approach and methodology, they can have the advantage of focusing on a well defined set of research activities with clear inputs, objectives and research plans. However, they often run into problems of attribution that can sometimes be addressed by examining a wider range of an individual's portfolio of research. The challenge for higher education institutions is to identify and catalogue impacts that are generated from a high volume of individual research activities from a number of very disparate disciplines. There are currently several well established tools and methodologies to capture academic impact, including citations and bibliometric information (Jarwal, Brion, & King, 2009). However, socio-economic impact has not yet been established as a mainstream outcome of research activity that can be captured and harnessed by institutions. Therefore, there are very fewer tools available and less understanding on how to identify, capture and evaluate systematically socio-economic impact on a wide scale. There could be a number of strategies for the identification of impact at an institutional level: socio-economic impact could be identified by the individual researcher, through local arrangements within research centres and groups, or through central research support facilities. Nevertheless, whilst researchers in some disciplines might be accustomed to seeing socio-economic impact through working closely with their users‟ communities, this would not be true for all researchers, research groups or disciplines. Therefore, the first step for the identification of socio-economic impact must be a programme of education to share and embed in the academic community definitions of socio-economic impact and how to recognise it. Central research support teams within institutions have a key role to play in supporting research (Shelley, 2010) and, therefore, to expand their responsibilities to include the identification and capturing of socio-economic impact. They can both support individual researchers and groups in identifying socio-economic impact, as well as systematically „scan‟ research activities and outputs, so as to be alerted to those activities that have a potential for creating socio-economic impact. Since, it is accepted that only a proportion of research activity will create significant socio-economic impact, institutions have to adopt strategies to focus on those that have characteristics that might lead to the adoption of their findings by policy makers, users‟ communities, or might influence social welfare and quality of life.

Defining socio-economic impact

One of the first important points when defining socio-economic impact is the distinction between „dissemination‟ and „impact‟. The report on the ESRC International Symposium Approaches to Assessing the Non-Academic Impact of Social Science Research (ESRC, 2005, p. 23) has noted, as one of its main conclusions, that while disseminating to a wide range of audiences is positive, impact has to be evidenced by the application of the research outcomes by the user or community. It also highlights that, "it cannot be assumed that the former [dissemination] naturally or inevitably leads to the latter [impact]”. Godin and Doré (2004) reaffirm the distinction between outputs and impacts where the latter are defined as changes in society directly related to the research activity. This was also a key element in Buxton and Hanney‟s distinction between Primary Outputs, Secondary Outputs, and Impacts or Final Outcomes (Buxton and Hanney, 1996). The Warry Report, commissioned by the UK Director General of Science and Innovation, asserts that, "an action or activity has an economic impact when it affects the welfare of consumers, the profits of firms and/or the revenue of government" (RCUK, 2006, p. 10). Yet, it goes further by adding to economic impacts that are easily quantifiable in terms of profit, low prices or higher revenues, those that are less quantifiable and affect the environment, public health and the quality of life. While the easily quantifiable economic impacts can be broadly defined and understood, those that affect welfare are ill-defined and highly subjective. There have been numerous reports and studies that have catalogued what might be considered as having impact on social welfare. Examples of these include: The British Academy‟s report lists five key functions of the arts, humanities and social sciences in contributing to social welfare and quality of life, including contributions to cultural and intellectual enrichment (British Academy, 2004). This concept is further developed by recognising, as socio-economic impacts, the following: influence on cultural performance; expertise to support the historic environment; supporting the sustainability of national identity, multicultural tolerance and interaction; fostering the debate that involves conflicting moralities, traditions and beliefs; and support for civic virtues and open government. The PA Consulting and SQW consulting report (PA Consulting/SQWconsulting, 2006), commissioned by Research Councils UK, groups impacts from funded research into four headings: development of human capital, business and commercial, policy, and quality of life. Quality of life is an especially difficult grouping to define and the report underlines how contributions to quality of life generate „social benefits over and above efficiency saving‟. Their description of quality of life benefits include: healthcare benefits for patients and carers; prediction of environmental impacts; social cohesion and enhanced security; and cultural advances including improvements in our understanding of the fundamental laws of nature.

Godin and Doré (2004) present a typology of categories of the impact of science on society (see table 2 below, where „culture‟ refers to how individuals understand the reality around them; „society‟ is concerned with behaviours and practices of individuals and groups; and „symbolic‟ refers to the gains in credibility of non-academic communities from being associated with academic research):

What emerges is that the definition of socio-economic impact is very broad and diverse and covers all aspects of social welfare. Under such broad definitions the challenge for institutions is to capture all socio-economic impact of their research activities and to be able to evidence the outcomes.

Institutional strategies for the management of the socio-economic impact of research

The challenges for institutions are multiple, but they focus around three main areas of activity: 1. How to support academics in the identification and tracking of socio-economic impact.

2. How to devise a system for academic managers to be able to systematically assess and compare the different socio-economic impacts generated by the research activity.

3. How to identify strategies that might enhance the achievement of high socio-economic impact.

Identification and tracking of socio-economic impact There are several models for illustrating the mechanisms or pathways from research to research outputs through to socio-economic impacts, some of which are related to specific disciplines or areas of research. The communality is that they all attempt to describe why and how impact was generated incorporating problems with time lag and attribution of impacts. The study of these pathways can provide a considerable advantage in understanding which characteristics to look for when determining the likelihood of a research activity producing socio-economic impact. An institutional plan for the collection of socio-economic impact should, therefore, firstly be concerned with the most effective way to target research activity and research areas that have a high probability of resulting in socio-economic impact. This can be done by concentrating on a number of characteristics (for example, active user involvement at the research design stage, or dissemination of research outcomes to knowledge brokers/lobby groups/professional associations) that can be identified by understanding the generation of socio-economic impact pathways and mechanisms. However, given that socio-economic impact must, by definition, arise in a non-academic community, the minimum requirement for any research activity is to have, on their path to impact, at least one 'contact-point' outside academia. This is schematically illustrated in figure 1 by two distinct pathways. These can be direct (Fig. 1a) as, for example, research that is developed in close collaboration with a particular user and, therefore, by definition in constant contact with the user. However, they can also be indirect (Fig. 1b) in cases where, for example, an algorithm developed in a discipline such as applied mathematics is then employed by engineering research to solve a particular industrial problem.

The diffusion of socio-economic impact needs to be considered also and might vary, as schematically illustrated in figure 2. Diffusion models reflect such variations while also noting timescale effects and strength of impact of an innovation (see for example the seminal works by Rogers, 1962, and Brown, 1981, and the recent major review by Greenhalgh, Robert, MacFarlane, Bate, & Kyriakidou, 2004). Nevertheless, the engagement of users (in the broadest sense), whether proactively or by chance, is an essential prerequisite for the generation of socio-economic impact. For example, the spread of research in a non-academic community can occur by a user spreading the research findings as „good practice‟ within its user community (Fig. 2a), or, by a research activity directly engaging with a number of user‟s within a community (Fig. 2b).

The first point of contact outside academia might not always be an end-user, but, an organisation or individual that can broadly be defined as a knowledge broker. As noted as long ago as 1983, knowledge brokerage roles can be structural, as through working parties or committees, or they can reside in individual roles such as those of liaison officers (Kogan and

Henkel, 1993; Kogan, Henkel, & Hanney, 2006). An organisation conducting a brokerage role might not directly benefit from the research activity, but can operate as a facilitator or intermediary between research activity and end-users. Examples of these organisations are learned societies, professional bodies and even the media. These groups tend to draw outcomes that might be useful for the community they represent from the academic knowledge pool. An example of these sorts of engagements concerns the research undertaken by one of the authors on copyright infringement in the Textile Industry (Dickson & Coles, 1998), where the research findings were disseminated in the form of non-academic reports and conference presentations not only directly to the user firms but also to professional bodies (such as Art & Design Colleges) and lobby groups (eg, ACID – Anti-Copying in Design) that represented such firms.

These representations of pathways to impact and the focus on user contact-points are powerful tools when systematically engaging with the academic community. The obstacles in trying to identify where research activity could have led to impact are greatly reduced by focusing attention on the most simple prerequisite of user engagement. This represents only the first step that indicates the dissemination of the research outcomes to non-academic environments. However, given that dissemination and engagement are not proxies for socio-economic impact, focus must immediately be directed to verify if the engagement activity has resulted in a change as described in Godin and Doré (2004) typologies of categories of the impact.

Systematic assessment and comparison of socio-economic impacts

Brunel pilot study framework

A pilot conducted at Brunel University (Scoble, Dickson, Fisher, & Hanney, 2009) revealed that a matrix type framework (Brunel Research Impact Device for Evaluation – BRIDE) can helpfully assist in assessing and comparing socio-economic impact. Building on an earlier use of two dimensions on which to evaluate the extent of the impact (Hanney, Davies, & Buxton, 1999), the 2009 pilot showed that using depth and spread as the two primary criteria to evaluate impact originating from research activity allowed for a systematic assessment of different types of identified impacts. The depth dimension indicates how transformational, to the end-user, the application of the research findings was, while the spread signifies how large a community or communities it has affected. The figure below illustrates the matrix and shows how, having four different levels for both depth and spread, the matrix can be shaded into areas that might be assessed as having comparable socio-economic impact.

While the level of depth (or change that the research activity has generated) and the spread (or how far it has permeated the user community) are based on subjective judgements, the placement on the matrix of the different cases of impact will be consistent relatively to each other.

The use of these two dimensions has also been defined as a way of assessing socio-economic impact in the HEFCE consultation document for the forthcoming Research Excellence Framework (HEFCE, 2009, p. 21). In the consultation document HEFCE describes the assessment of socio-economic impact based on two criteria:

・“how significant or transformative the impacts have been”

・“their reach (how widely the impacts have been felt)”.

Thus the terms 'depth' and 'spread', as used in the BRIDE matrix, can be easily interchanged with 'significance' and 'reach'. The use of the BRIDE assessment matrix can also be extended to include a third dimension that could capture the types and nature of the benefits and, therefore, help develop a narrative of the overall socio-economic impact. The main objective of this type of framework is not to determine the absolute value of the socio-economic impact, but mainly to identify, relative to each other, which cases might need further investigation or where there is potential to increase their socio-economic impact. Once cases are identified through the BRIDE assessment matrix, there is scope to develop strategies to encourage and enhance the generation of impact from the research activities.

Strategies for the enhancement of high socio-economic impact

An analysis of any case study in any of BRIDE‟s cells will show that in order to enhance the socio-economic impact of the research activity there are two main strategies to pursue. Movement from one cell to another is a combination or vector of two factors, as illustrated in the figure below: increase in depth/significance and/or increase in spread/reach.

Depth/significance is a measure of how profoundly the results and outcomes of a research activity has affected end-users and communities either through transforming businesses and industrial process, changing attitudes, improving the quality of life and other elements as illustrated in the typology of categories of impact by Godin and Doré (2004). Therefore, it can be inferred that improvement of depth/significance can be achieved by either further indepth research or by identifying end-users and communities where the outcomes of the research activity might be more beneficial. Spread/reach is a measure of uptake of the research activity‟s outcomes. Any end-user or community can only be affected if they have been, either directly or indirectly, exposed to the research activity findings. It is, therefore, strategically beneficial to have a coordinating plan to maximise dissemination and active engagement of end-users and communities that might benefit from the research activity. As a result, any strategy to improve the possibility of enhancing socio-economic impact of a research activity must be based on building the depth/significance through further quality research and developing the spread/reach through focused dissemination and engagement of non-academic communities. A typical realistic example may help to illustrate the proposed two-pronged strategy. Consider a short pilot research project that is undertaken to identify critical factors in the survival of local small-scale theatres, funded by a regional authority. Thus it would be located in the bottom left-hand corner of the BRIDE assessment matrix, as indicated by the letter „A‟ in fig 6 below, since the research is fairly superficial, being only a pilot study, and of limited spread due to its regional nature. By doing more extensive research, with funding from say a national UK body (eg, ESRC or The Leverhulme Trust) or even from the European Union, so that wider and deeper coverage of relevant issues can be addressed, it is expected that more substantial findings and better understanding of how small theatre survive in differing socio-economic settings would be generated, leading perhaps to position „B‟ in the matrix. It is possible that such extended research might develop theory on the economics of drama or on entrepreneurial identity in creative industries. New theoretical insights might then lead to new national or European policy or support initiatives for regional theatre, thus relocating the impact position to say „C‟. Alternatively, the pilot study might generate findings that can be reproduced as training materials for new theatre founders and managers, so the impact moves from „A‟ to „D‟ as the training material gets adopted by other theatres in other regions. If that proves successful in improving the survival rate of small theatres, then perhaps the training material will be approved and adopted by some major national body such as the Arts Council in the UK whose influence is widespread and may even extend beyond the UK. The effect would be to move the impact further to the right, say to „E‟ in the matrix in fig 6. Of course, both pathways, from „A‟ to „C‟ and from „A‟ to „E‟, could occur simultaneously. Indeed, such complementarity is desirable and, no doubt, successful ambitious researchers would seek out any combination of opportunities to increase impact.

Conclusions While the emerging agenda for the evaluation of socio-economic impact is challenging, coherent institutional strategies can be planned based on the understanding of the fundamental pathways from research activity to impact and by establishing frameworks based on key evaluative criteria. The key to success in embedding socio-economic impact in the institutional research capital, given limited resources, might rely on the ability to focus researchers on the essential minimum requirements for the generation of impact and to guide them through the pathways to socio-economic impact. This approach of tracing a simplistic roadmap helps individuals identify possible occurrences of socio-economic impact that might otherwise be disguised by a lack of understanding of what needs to be sought. Once this type of approach is fully understood and implemented, tracing of socio-economic impact should become more apparent and embedded as one of the many activities for the identification of the institution‟s research capital. The framework for the evaluation of socio-economic impact has to be a flexible and adaptable tool that will be most beneficial for the identification of strategies to enhance the possibilities of creating high-value socio-economic impact. Criteria such as depth/significance and spread/reach are essential to determine, on a basic scale, the relative value of impact. However, further categorisation of the framework to enable the evaluation by type of socio-economic impact can provide a rich picture of the overall benefits that the institutional research capital has on non-academic communities. The expansion of institutional research capital by the inclusion of the socio-economic impact is still at an early stage. Different approaches and frameworks will be developed for capturing and evaluating socio-economic impact. In addition, these developments will be coupled with an increased understanding of disciplinary-specific pathways from research to impact and with a more engaged academic community alert to the need to follow the outcomes of the research beyond the academic domain. The growth in understanding and the development of new knowledge and new models will, inevitably, see the consolidation of systems and strategies throughout the academic sector.

It could be said that Spanish Communication research a decade ago was not especially prodigious when it came to taking Communication research itself as an object of study and to reflecting on the knowledge interests and epistemological orientations thereof, or on the results obtained therefrom (Martínez Nicolás, 2008). Since then, however, efforts made in this sphere have begun to put us in an optimal position – or at least a more reliable one than before – to reconstruct paths and propose diagnoses of the historical evolution of Communication research in Spain. Back then, we had a number of valuable contributions on this issue, from the pioneering works of Moragas (1981) and Gifreu (1988 and 1989) on the history of the field of Communication, to the subsequent contributions by Jones in the 1990s and early 2000s (Jones, 1994, 1997, 1998a and 2000). However, it has only been in the past ten years, since the mid-2000s, that interest in what we now call meta-research has taken hold with an intensity that is probably without parallel in other disciplines in Spain, or indeed to what is happening in other countries with similar scientific cultures. Focusing mainly on academic production in specialised journals, and to a lesser extent on doctoral theses, the effort put into this task of recognition should not lead to complacency because certain weaknesses and limitations that we noted 10 years ago continue to exist. To begin with, works aimed at the critique of research are still few and far between. Such works should necessarily be limited to a particular specialised domain (a particular sub-discipline or an object of study, a theory, a methodological perspective, etc.) to ensure that the validity, thoroughness, novelty or originality of our contributions can be properly assessed. Along these lines, the collection of essays on Communication research in Catalonia, edited by Berrio, remains an essential work of reference. Since it was published more than 20 years ago, there has hardly been any follow-on work (Berrio, 1997). Lacking this somewhat more sophisticated critical orientation, the recent rise in Communication meta-research in Spain seems to be too concerned with bibliometrics and content analysis of more or less stringent samples. Generally in the form of scientific articles or conference presentations, such research only scratches the surface or shows the most superficial trends, if you will, of the research we are actually conducting. Furthermore, and in a fairly generalised way, such meta-research usually falls into the trap of mere descriptivism. Consequently, it provides us with a certain representation of the state of research (objects of study, methods, techniques, authorship regimes, epistemic networks, contributing universities, etc.), but it does not manage to explain why research is actually in that state, thus overlooking a variety of different factors (socio-historical, institutional, epistemological, science policy-related, etc.) that have an impact on it at any given time. With its scope and explanatory power curtailed, the capacity of meta-research to serve as a stimulus for research practice is diminished. Indeed, this is contrary to the expectation that research practice should identify the shortcomings of meta-research, suggest new ways to develop it and elucidate the determining factors thereof. Furthermore, such practice can only be carried out inasmuch as it is formulated as a critique of knowledge and approached from the sociology of science perspective, which addresses the social, historical and institutional contexts within which scientific activity inevitably falls (Kuhn, 1962; Merton; 1973; Bourdieu, 1984 and 2001). Be that as it may, the vast amount of empirical evidence gathered in the past decade puts us, as we said, in an optimal position to give an overview of the evolution of Communication research in Spain over the past 25 or 30 years. It enables us to make headway towards reconstructing a general history of the field by building on the works by Moragas (1981) and Gifreu (1989). And that overview not only shows that the institutional framework supporting Communication research in Spain has changed radically over the past three decades (Martínez Nicolás, 2008; Saperas, 2016), but also that the process has affected the structure and practices of the academic community, and particularly the characteristics and general direction of scientific production in this sphere. Successive provisions relating to general science policy in Spain have undoubtedly played a decisive role in bringing about such change. Strengthening the culture of evaluation has especially done so, since it led to the creation of Spain’s National Agency for Quality Assessment and Accreditation (ANECA, as abbreviated in Spanish) in 2002, and to the implementation of the Academia programme for university teaching staff accreditation in 2008. Indeed, their role has been so decisive that, as we shall endeavour to demonstrate in this work, the various indicators available to us on the evolution of Spanish Communication research over the past 30 years show that there was a turning point in the mid-2000s, a time when this disciplinary field seemed to enter a new phase characterised by a scientific culture that was very different from the previous one. However, the factors explaining that step change go beyond the often mentioned “ANECA effect” (Soriano, 2008). While useful for fostering reflection on the field – to the extent that the take-off of Communication meta-research in Spain can be considered driven by that “effect” – it has ended up becoming a kind of flatus vocis used to expeditiously get rid of references to the “context” within which Communication research in Spain has been conducted over the past decade. Without denying its relevance, the impact of State policy decisions condensed in that “ANECA effect” does not consider the array of factors determining scientific activity in this sphere, not even in the past 10 years. Moreover, it remains to be ascertained, in detail and with adequate empirical support, in what sense an “effect” – one that had often become the source of many of the “defects” afflicting Spanish Communication research – would have been felt. Although we now have some works that address this issue (Masip, 2011; Goyanes, 2015 and 2017; Lacasa, 2017; Martínez Nicolás, Saperas & Carrasco, 2017; Soriano, 2017), the systematic review of the available evidence that we propose in this work may help to reconstruct the complexity of the Communication research system in Spain over the past 30 years. This necessarily requires us to address the interaction between the institutional context, the structure and practices of the academic community, and scientific production in that period.

2. Institutional changes affecting Communication research

The institutionalisation of Communication research in Spain began with the creation of Information Sciences faculties at the Complutense University of Madrid and the Autonomous University of Barcelona in 1971. A few months later, in April 1972, Journalism studies taught at the University of Navarra – governed by the Catholic Prelature Opus Dei – were recognised at university level (Decrees 2478/1971 and 891/1972, respectively). This process culminated in the creation of the faculty of Information Sciences at the University of the Basque Country in October 1981 (Royal Decree 2344/1981). Communication Studies were then organised into a single degree in Information Sciences, with three branches or specialisations in Journalism, Audio and Visual Image Sciences, and Advertising (Decree 2478/1971). The sectoral structure adhered to in order to train Communication professionals at the former Official Schools (of Journalism, Radio and Television, Cinematography, and Advertising) was transferred to the university sphere and, as a consequence of the new faculties taking on their functions, those Schools gradually became defunct. But, unlike them, those new faculties constituted an institutional framework conducive to starting to forge a scientific tradition of Communication research in Spain. This was so because, among other reasons, they generated a structure of opportunity for academic professionalisation, whose mechanisms of access to and promotion during an academic career (doctoral theses, competitions to fill vacancies, etc.) required a decisive commitment to scientific activity (Martínez Nicolás, 2008). However, those access and promotion mechanisms were still too lax, and they were often dependent on struggles between groups of academics to form clientelist networks instead of on the strict application of meritocratic criteria. While certain aspects of such inertia persisted – and still persist today – the fact is that the institutional context within which Communication research in Spain was conducted from the mid-1980s was substantially undermined as a result.

2.1. Increase in the number of universities offering Communication studies

The pace of the institutionalisation process that began in the 1970s became more intense from the 1990s, with the boom in Communication studies in Spain. In the mid-1980s, the four faculties in existence at that time together had 5,000 students and around 500 lecturers. No more than 10 years later, those figures had quadrupled. By that time, 20 universities had begun to offer such studies, and between them they had 20,000 students and some 2,000 lecturers (Jones, 1998a). That expansion in offerings continued to grow exponentially. A decade later, in the mid-2000s, 44 universities offered such degrees and there were around 3,000 lecturers (ANECA, 2005; Abuín, 2010). The most reliable calculations for the recent period have estimated that, in 2015, there were 54 Spanish universities offering Communication programmes, on which there were 45,000 students being taught by some 4,500 lecturers (Saperas, 2016). Thus, regardless of the reference parameter taken – universities, students or lecturers – the data show that the institutional volume of Communication studies in Spain has multiplied tenfold over the past three decades.

Possible reasons for this inflationary process are the attractiveness of Communication-related professions and the large demand for such professionals generated by the radical transformation of the Spanish Communication system in this period. In the 1980s, the lifting of the iron-grip control exerted until then by the Francoist dictatorship, and Spain’s political alignment with Western democracies – sanctioned by its accession to the European Union in 1986 – favoured the liberalisation and professionalisation of the sector (press, radio and film in particular). That push to liberalise and professionalise the sector continued until the early 1990s, when the public television monopoly was broken up and private television outlets began operating in Spain, thus fostering growth in the Audiovisual and Advertising markets. The expansion and dynamism of the Communication system has begun to drop off in recent years as a result of the advent of the digital revolution in the early 2000s. That revolution is at the very heart of the profound social, political, economic and cultural changes that advanced contemporary societies are experiencing. Driven by a variety of circumstances at any given time, but always in an expansive direction, Communication studies have gradually taken root among the university education options with the highest demand and greatest potential for employment in Spain in the past 30 years. The exponential multiplication of offerings in this period may therefore have been due, to a large extent, to the decisive incorporation of newly created private universities into this educational sphere (Moragas, 2005). Such universities accounted for a third (19 out of 54) of the registered Spanish centres offering Communication degrees in 2015 (Saperas, 2016, p. 37). In that same year, there were 34 registered private universities in Spain (Simancas & García López, 2016, p. 181). Consequently, nearly 60% of those universities offered Communication degrees, thus giving a clear indication of the attractiveness that such studies have gained in private higher education offerings in the past 15 to 20 years.

2.2. Establishment of distinct Communication degrees

In the early 1990s, the old-style Degree in Communication Sciences, with mentions of specialisation in Journalism, Audio and Visual Image Sciences, or Advertising was split into three distinct degrees in Advertising and Public Relations, Audiovisual Communication, and Journalism (Royal Decrees 1386/1991, 1427/1991 and 1428/1991, respectively), with curricula that had also been updated. This option had probably been prefigured in the decision to segment the disciplinary field of Communication into two distinct “knowledge areas”: “Journalism” and “Audiovisual Communication, and Advertising”. That decision was set out in the provisions emanating from Spain’s University Reform Law (1983) to regulate access to university teaching staff faculties (Royal Decree 1888/1984) and, consequently, the most specific nucleus of teaching staff in Communication Sciences faculties compulsorily came under one of those two areas from then on. And, after the 1991 reform, the term Communication Sciences became generalised. Be that as it may, the fact is that the impact that the establishment of distinct degrees would have on the Communication research system in Spain over the following quarter century might not have been assessed properly. Even after the implementation of the three specialisations, the first faculties of Information Sciences – a label that had already become highly significant – were centres whose basic aim was to train journalists, thus generating a scientific community whose attention was focused on that media sphere. Furthermore, this was happening shortly after Franco’s death, the demise of his dictatorship, and Spain’s transition towards democracy, when renewed value was placed on the social role of Journalism and the Media, hence the interest in and urgency of their analysis (Martínez Nicolás, Saperas & Carrasco, 2017). In this respect, it was symptomatic that, when it came to segmenting the disciplinary field into “knowledge areas”, a decision was taken to single out “Journalism” and group “the rest” into a second area. By doing so, it highlighted the somewhat subsidiary position that the spheres of Advertising and Audiovisual Communication occupied in the academic mindset at that time. And that is without taking Public Relations into account, which had been officially missing until the 1991 degree reform. The implementation of distinct degrees had the symbolic significance of putting the field’s various specialisations on an equal footing in terms of institutional recognition, hence the replacement of the name Information Sciences with Communication Sciences in faculties from that moment on. The immediate academic effect of that decision was to strengthen university professionalisation and scientific activity in certain specialised niche areas (Advertising, Public Relations and Audiovisual Communication), which had been overlooked by Spanish Communication research until then.

2.3. Deployment and consolidation of a culture of evaluation

The impetus to modernise, embodied in the University Reform Law passed in August 1983 (Organic Law 11/1983), was aimed mainly and decisively at dragging Spanish universities out of the long-standing state of scientific anaemia that had hampered Spain’s ability to take its place among advanced industrialised societies. The preamble was very clear in this respect: “Even if it were only to boost development, mentality and the scientific spirit in Spain, university reform would be justified.” And that was the leitmotiv of the University Reform Law. The measures proposed with that objective in mind included, among other provisions, a rearrangement of university structures, with the creation of departments, unto which it entrusted the organisation and coordination of scientific and teaching activities (Article 8); the reform of doctoral programmes, which should be aimed at “student specialisation and training in research techniques” (Article 31); and, in particular, the requirement that teaching staff should be subject to “an evaluation of teaching and scientific performance”, which would be taken into consideration in competitions to access tenured teaching staff faculties (Article 45). This was the starting point for a culture of evaluation of research activity that, later on, would revolve around the concept of scientific productivity, whose assessment criteria would have a significant impact on the general direction of scientific research in Spain from that moment on. In the development of that culture of evaluation, at least three stages can be distinguished: implementation (1983-1989), strengthening (2001) and generalisation (2008). It has evolved in a way that not only makes the criteria tougher and the thoroughness of their application stricter, but also and mainly it has made scientific performance a determining factor for access to, consolidation of and promotion during the researchers’ professional careers. In the implementation stage, it was characteristic to consider scientific performance as a mere incentive to conduct research. In fact, that notion began to circulate as a consequence of a seemingly minor provision in terms of scientific policy about teaching staff’s pay (Royal Decree 1086/1989). That provision established the six-year rewards, which are well-known among Spanish academics. Interested parties have to apply for them voluntarily, and the body in charge of granting them was a newly created one: Spain’s National Commission for the Evaluation of Research Activity (CNEAI, as abbreviated in Spanish). That state of affairs continued throughout the 1990s and until the Law replacing the University Reform Law was passed in December 2001. The new one, the Organic Law on Universities (Organic Law 6/2001), was a veritable turning point for the consolidation of that culture of evaluation in Spain. Indeed, Title IX of the Organic Law on Universities established two ways to access the many categories of university teaching staff positions, one reserved for tenured teaching staff faculties (professors and senior lecturers) through a system of “national habilitation” that entitles them to take part in access competitions; and another for academically more demanding categories of non-tenured teaching staff (assistant doctor/lecturers and contracted doctor/lecturers). Imposed on the latter is an obligation to have obtained a positive evaluation of previous, “primarily research” activity in the case of contracted doctor/lecturers (Article 52) from another newly created body: Spain’s National Agency for Quality Assessment and Accreditation (ANECA, as abbreviated in Spanish). That body was formally constituted in July 2002, when, alongside regional agencies created afterwards, it began to carry out the functions conferred on it by the Organic Law on Universities. The novel feature introduced by the Organic Law on Universities was, therefore, that scientific performance was no longer an incentive but instead an inescapable requirement for developing an academic career, thus reinforcing the culture of evaluation implemented a decade earlier. And, from April 2007, those conditions were immediately generalised across the teaching staff community by the Organic Law Modifying the Organic Law on Universities (Organic Law 4/2007), which also extended the requirement for prior accreditation exclusively by ANECA to tenured university teaching staff faculties ANECA (Modification of Article 57 of the Organic Law on Universities). In January 2008, little more than half a year later, ANECA implemented the Academia programme for university teaching staff accreditation, which made “research activity ” a core criterion for obtaining the positive assessment required to be able to get access to and professional promotion at Spanish universities. Thus, for example, out of a maximum of 100, the minimum score for accreditation as a Tenured Senior Lecturer was set at 65 points, 35 of which (nearly 54%) had to correspond to the research merits provided, and the remainder to teaching activity and university management. The cornerstone of the deployment of this culture of evaluation, and the undeniable cause of numerous objections from the sphere of Social Sciences (Jiménez, De Moya & Delgado, 2003; Perceval & Fornieles, 2008; Soriano, 2008; Ruiz Pérez, Delgado & Jiménez, 2010; Quirós, 2016; Soriano, 2017), was the definition, by the bodies in charge of implementing it, of what should be understood as “research activity” and, above all, of what the most suitable criteria for assessing the quality of a scientific track record should be. For the purpose of granting six-year rewards, CNEAI was charged with that task. From the outset, it decided to establish “quantifiable and verifiable criteria” in order to increase “scientific productivity” so that it could be “detected in critical international databases”, in the words of Ana Crespo, the Commission’s general coordinator from 2005 to 2008 (cited in Soriano, 2008). And that approach was adopted nigh on unequivocally by ANECA in the procedures for teaching staff accreditation. CNEAI’s relentless job of refining what counted as “research activity”, as well as the quality standards attached to it, was reconstructed in detail up to 2009 by Ruiz Pérez, Delgado and Jiménez (2010). And, revised to the current time, it reveals a kind of pendulum movement of tightening and loosening in their approach. Currently, it is in a tightening phase. In short, the most significant milestones in this process were the practical reduction of scientific activity to “publications” at the expense of any other type of research product or task (conference presentations, report elaboration, doctoral theses supervision, project or research group management, etc.); and, above all, the preference given to articles published in academic journals. Indeed, the journals’ impact factor was introduced as a quality criterion in 1996, with an explicit mention of those included in the Journal Citation Reports (JCR), a citation index managed by a company dedicated to the publishing business. That decision had an immediate effect on the sphere of Social Sciences because, as highlighted by Ruiz Pérez, Delgado and Jiménez (2010, p. 906), “when the criteria for ‘hard sciences’ are applied to some extent to ‘other forms of knowledge’, the success rates [in the evaluation of research activity] fall considerably”. Thus, between 1989 and 2004, only 59% of the six-year rewards applied for in the Social Sciences field were granted, the lowest alongside Economic and Business Sciences. In the 2005-2008 period, CNEAI’s successive provisions sought to “soften the previously pronounced JCR- centrism” (Ruiz Pérez, Delgado & Jiménez, 2010, p. 907) by recognising other Spanish and international indices as being indicative of the quality of scientific journals (Scopus, Latindex, In-RECS, DICE-CINDOC, etc.). But, for Social Sciences, such “softening” ended with the 2016 call for six-year rewards applications. Whereas the 2015 call stated that publications in JCR journals (point 3a), Scopus journals (point 3b) and other journals at the assessors’ discretion (point 3c) would be rated more highly, the 2016 call reinstated the privileged status of JCR; journals in high positions in Scopus would be accepted “without them necessarily being rated equally”; and, lastly, it stated that the remainder “could also be rated, though never in the same way”. The language used by CNEAI was certainly and surprisingly expeditious, once again shifting the pendulum to the “tightening” position. As we said, ANECA nigh on unequivocally adopted the criteria that CNEAI had established for university teaching staff accreditation, and even followed the latter’s pendulum movement. Implemented in 2008, the evaluation of research activity within the framework of the Academia programme gave precedence to publications in indexed journals, described generically as “catalogues like the Journal Citation Reports or equivalent”. Such laxity began to be redressed with the provisions specifying the new procedure for university accreditation approved in 2015 (Royal Decree 415/2015), which, for the field of Social Sciences, ordered scientific journals into four levels, starting with those in the first and second quartiles of JCR and the first quartile of Scopus. The provisions also required a minimum number of published works in these to obtain a positive assessment (for example, for the Tenured Senior Lecturer category, 10 articles in the first level to attain the highest grade (A), and four to obtain the minimum grade (B) actually required). As a result of the academic community’s response, the accreditation system had become practically paralysed until those criteria were reviewed, but it resumed in 2017 with new indications about the evaluable merits1, which continued along the same lines as before, that is, of strongly rewarding scientific activity giving rise to articles in journals included in the international rankings calculated by taking into account their impact factor (i.e., the number of citations of the articles they publish). With the new tools at their disposal for independently disseminating their works in the web environment, a researcher’s prestige might potentially depend less and less on the journals in which her or his work is published, and more on her or his own “digital reputation” (Delgado, 2017). Be that as it may, the assessment of scientific activity quality seems forever doomed to be dependent on rankings, citations and followers. We may be measuring their “impact”, but we are certainly not appreciating their quality or influence: there are no algorithms for that2. In sum, the deployment of a culture of evaluation based on performance and productivity parameters thus defined ended up determining, and more forcefully from a decade ago, the general system of scientific research in Spain, and it obviously affected the disciplinary field of Communication too.

2.4. Recognition of Communication in the National RD&I Plan

The act of boosting “mentality and the scientific spirit in Spain”, which served as a catalyst for the 1983 university reform, began to be firmed up with the passing of the Promotion and General Coordination of Scientific and Technical Research Law in 1986 (Law 13/1986). Known as the Science Law since then, it not only corroborated the long-standing backwardness of scientific activity in Spain, but also diagnosed the woes that beset it, with a number of remedies being proposed. The presentation of motives for the Law was also very clear in this respect, thus maintaining the drive that was characteristic of legislation in this field at that time, it stated: “In Spain, scientific research and technological development have traditionally taken place in a climate of atony and lack of social stimuli, [and] absence of instruments to guarantee the public authorities’ effective intervention in the programming and coordination of the scarce resources available [...]. It is not surprising, therefore, that the Spanish contribution to scientific and technological progress has generally been scant and unfitting of the place corresponding to us in other areas”. To rectify that situation, and among other provisions, the Law provided the Spanish scientific system with an instrument: the National Plan for Scientific Research and Technological Development. That plan proved fundamental to sustaining an active policy for stimulating research activity in Spain. The National Plan is organised into specific programmes covering the various spheres of the science and technology system (researcher training, mobility and integration, infrastructures, projects, institutional strengthening, etc.) in successive four-year plans, from the period 1988-1991 to 2017-20203. Of these programmes, the ones with the most decisive and sustained impact on scientific activity are probably those aimed at research project funding, on which the development of researchers’ ordinary work under more or less optimal conditions has depended considerably since then. After several restructuring efforts, the management of such projects has ended up being organised by large departments (Humanities and Social Sciences, Life Sciences, Environment and Natural Resources, etc.) that include the different areas and sub-areas of knowledge4. A characteristic of such management was, however, that the funding of projects submitted to annual competitive calls for applications would be resolved by a strict system of peer evaluation, in which the different panels of experts ruled on the quality of the projects and the amount of funding needed to carry them out in a maximum period of three or four years. The sphere of Communication’s position in that organisational structure was certainly precarious over the more than 20 years between the first and sixth National Plans, in which the project applications generally targeted one of the sub-areas of knowledge in the field of Social Sciences. Those sub-areas were usually Sociology and Political Science, hence the identification codes SOCI and CPOL (as abbreviated in Spanish) applied to those that were successful in 2010. These were consequently assessed by experts in those sub-areas. That generated a more or less justified sense of grievance among researchers assigned to Communication faculties, who complained that the relatively little success of their proposals was due to the fact that the people assessing the projects in this sphere were not “specialists in Communication”. A sense of grievance, we insist, that was more or less justified because it may also have been the case that those “Communication” projects did not reach the required level of quality due to their inherent shortcomings (Martínez Nicolás, 2008). Whatever the case may be, after persistent and ultimately successful demands by the scientific community, that issue of Communication’s precariousness began to be rectified from the 2010 call 3 Since its implementation, there have been eight successive four-year plans: 1988-1991, 1992-1995, 1996-1999, 2000-2003, 2004-2007, 2008-2011, 2013-2016 and 2017-2020. National Plan for Scientific Research and Technological Development (National R&D Plan) was the original title established under the Science Law, which was changed to National Plan for Scientific Research, Development and Technological Innovation (National RD&I Plan ) from the fourth plan drawn up for the 2000-2003 period. After the Science, Technology and Innovation Law had been passed in 2011 (Law 14/2011), which replaced the previous one dating from 1986, the State Plan for Scientific and Technical Research and Innovation (the current State RD&I Plan) was established, whose first four-plan corresponded to the 2013-2016 period. The replacement of the previous National Plan with the current State Plan to bring it into conformity with the 2011 Science Law meant that the 2012 call for applications was still governed by the provisions of the sixth plan (2008-2011). The State Research Agency provided for under that 2011 Law was not created until 2015 (Royal Decree 1067/2015). The delay in it becoming fully operational (early 2017) explains why the last call for applications (2017) for research project funding was still aligned with the directives of the 2013-2016 plan, because the eighth and current one (for the 2017-2020 period) was published in December 2017, once the call for applications for that year had closed. 4 In the case of Humanities and Social Sciences, for example, those of Law, Economics, Education, Psychology, History and Social Sciences, among others; with the sub-areas of Sociology and Political Science coming under Social Sciences.

for applications within the framework of the sixth National R&D Plan, when the Social Sciences area was reorganised into four sub-areas of knowledge: Sociology, Political Science, Geography and Communication. Since then, they have also had specific “Communication” coordinators and experts to assess the submitted projects. Indeed, the recognition of Communication for these purposes had an immediate impact on the number of funded projects in the area from 2010. The demands relating to research activity imposed by the Academia programme for university teaching staff accreditation probably contributed to that too. In this respect, the data are fragmentary and uncertain, especially for the period prior to 2010, precisely because the sphere was spread across different areas of the National Plan. Be that as it may, between 2004 and 20095, a mean of 14 Communication projects per call for applications were successful (a minimum of eight in 2005 and a maximum of 21 in 2009). Then, in the eight calls for applications issued between 2010 and 2017, the mean rose to 25 funded projects per call, reaching a maximum of 32 in 2013. As we said, that boost can probably be traced back to the implementation of the Academia programme in 2008, with the significant increase from 12 projects approved in 2008 to 21 in 2009, the year when the research pressure exerted by the accreditation procedure began to be felt. However, it was reinforced by a decision that, in the period considered here (2004-2017), managed to practically double the number of funded Communication projects per call for applications: the recognition of Communication as a management area in the State RD&I Plan. It was a science policy measure that, as we shall see, would take Communication research in Spain in directions that researchers had been reticent to take.

3. Changes in the structure and practices of the scientific community

The changes that the institutional framework supporting Spanish Communication research has undergone over the past 30 years have left their mark on the structure and research practices of the scientific community committed to the development of this disciplinary field. Its exponential growth and acquired internal diversity; the gradual incorporation of women into Communication research; the intensity of the secondary institutionalisation process led by scientific associations, research centres and groups, and specialised journals and editorial collections; and the consolidation of its presence on international circuits of knowledge production and dissemination (publications, conferences, research projects, etc.) are some of the most outstanding features of the changes affecting the scientific community over the past 30 years.

3.1. Growth and internal diversity of the scientific community

The boom in Communication studies since the 1990s has led to a ten-fold increase in the number of Spanish universities that include Communication in their programme offerings, rising from the four pioneering universities to 54 universities in 2015. The number of teaching and research faculty members in this field has also increased by a similar rate. According to the most reliable data, the figure is around 4,500 members, nearly 10 times more than 30 years ago. The expansion of Communication studies has therefore led to a massive incorporation of lecturers on whom academic professionalisation expectations are placed, which requires – and particularly demandingly so since the implementation of the university accreditation system – them to develop a sustained research track record that is subject to external assessment.

A good indicator of such growth of the scientific community is the evolution of doctoral theses defended in Communication faculties since their creation (López Escobar & Martín Algarra, 2017, p. 97-98), totalling 250 between 1979 and 1989; 569 from 1990 to 1998; and 948 in the period 2007-20136. Although the time periods are not equal, a geometric progression is clear to see, in which the number of theses submitted and, therefore, the number of doctors doubles from one period to the next, reaching a mean of 135 new doctors in Communication Sciences per year in the most recent period (2007-2013). And while some of the members of that qualified staff were foreign graduates, mainly Latin Americans, who would go on to pursue their academic careers in their respective countries, the data are revealing enough to show the exponential growth of the scientific community and, above all, the pressure being exerted on a university access and promotion system linked, like never before, to research productivity. As already mentioned, the first Spanish Information Sciences faculties focused on Journalism teaching and research, and relegated the remainder of the field’s specialisations to a subsidiary position. However, the latter gained definitive institutional recognition with the implementation of three distinct degrees in Journalism, Audiovisual Communication, and Advertising and Public Relations in 1991. That decision had immediate academic effects, strengthening teaching activity and scientific research in domains that had traditionally been neglected by Spanish faculties. Moreover, those faculties would be responsible for the organisation and provision of knowledge for specific degrees and postgraduate programmes from then on. The data on the evolution of doctoral activity – i.e., on researcher training – in these spheres are enlightening: 60% of theses on Public Relations submitted to Spanish universities between 1965 and 2004 were defended between 1995 and 2004, that is, by the first groups of graduates of the degree in Advertising and Public Relations (Castillo & Xifra, 2006). Something similar happened in another specialisation linked to that degree, Advertising, in which 72% of the 260 doctoral theses recorded between 1974 and 2010 were defended from 1996 onwards (Marcos, Martínez Pestaña & Blasco, 2012). Regarding Audiovisual Communication, the data compiled by Repiso, Torres and Delgado (2011a) showed spectacular growth in the number of doctoral works on Radio between 1998 and 2007. Sixty-five theses were submitted in those 10 years alone, compared to the 52 submitted in the two previous decades (1978-1997). And the same thing occurred with those dedicated to Television (Repiso, Torres & Delgado, 2011b); nearly 70% of the theses submitted to Spanish faculties from 1976 were defended in that same period (1998-2007). Everything therefore seems to indicate that these spheres expanded considerably as soon as university degrees specific to them became available, thus contributing to a gradual diversification of knowledge interests to which the Spanish scientific community involved in Communication research attended.

3.2. Incorporation of women into the scientific community

In a field traditionally dominated by men, not enough attention has yet been paid to the incorporation of women into the scientific community. To begin with, we do not have at our disposal reliable figures about the distribution by gender of teaching staff assigned to degrees in Communication at Spanish universities. As a result, whatever we say about it must be based on the data provided by an analysis of the authorship of doctoral theses defended at Communication faculties, and of works published in specialised scientific journals. The latter are a good indicator of the characteristics of a scientific community that is really practising – i.e., one that contributes, by publishing, to knowledge production. However, they are usually circumstantial, referring either to specific moments in time (Castillo & Carretón, 2010; Escribà & Cortiñas, 2013) or to specific spheres (Piñeiro, 2016; Baladrón, Manchado & Correyero, 2017). Above all, they do not offer a historical perspective that would otherwise enable us to assess the process of the incorporation of women into Communication research. The data provided by those works showed a reasonably balanced distribution of their authorship by gender. That was what Castillo and Carretón (2010, p. 310) found, although the data referred exclusively to works published by a sample of journals in 2008. Escribà and Cortiñas (2013, p. 38) expanded the timeframe to the 2007-2011 period for works published in seven Spanish journals, and they found that men accounted for 60% and women for 40% of all authorships. The results obtained by Piñeiro (2016, pp. 40-41) showed a significant increase in the presence of women in Radio research between 1980 and 2013. However, the time periods they established were so long (1980-1994 and 1995-2013) that a clear progression could not be gleaned from them, particularly for the most recent period. Nevertheless, whereas women signed just 20% of the published works (7 out of 35) in the 1980-1994 period, that figure rose to a little over 50% between 1995 and 2013 (189 out of 367). From these data, that author inferred an “incipient feminisation of Radio research” in Spain (Piñeiro, 2016, p. 41). Baladrón, Manchado and Correyero (2017, p. 10) analysed the authorship of Advertising research between 1980 and 2015, with an overall distribution for the entire period of 50.3% men and 49.7% women; regrettably, they did not disaggregate the data by time period. More revealing still, due to its exhaustiveness, is the picture that can obtained from the evolution of doctoral thesis authorship. The data recorded by Jones, Baró, Landa and Ontalba (2000, p. 23) showed that theses produced by women accounted for a little over 10% of the total in the 1970s (9 out of 77). The proportion rose to around 30% in the 1980s (140 out of 460) and stagnated at a similar percentage in 1990-1998 period (37%, 368 out of 993). However, in the 2000s, that situation seemed to have changed significantly and, between 2007 and 2013, the 977 defended doctoral works were distributed equally between men and women (49.6% and 50.4%, respectively)7. For the time period between the two mentioned above (1999-2006), the specific case of Public Relations theses stands out in particular, since 56% of them were defended by women in the decade spanning 1996-2005 (Castillo & Xifra, 2006, p. 150). The balanced gender distribution in Spanish Communication research therefore seems to be a recent phenomenon, probably from the 2000s onwards. Be that as it may, this situation of balance does not yet reflect the “feminisation” of Communication studies in Spain. Although we do not have at our disposal historical data, or even data for the set of degrees, the explorations made so far are very clear in this respect. In the 1999/2000 academic year, around 70% of graduates of the degrees offered at that time were women (ANECA, 2005, p. 149); and nearly 15 years later, in the 2012/2013 academic year, the percentage of graduates in Journalism was practically the same (Rivero, Meso & Peña, 2015). Thus, although the presence of men and women in the scientific community is balanced, the position of the former in university teaching and research seems to be higher than their relative weight among Communication graduates.

3.3. Secondary institutionalisation of communication research

If the boom in Communication studies from the 1990s led to the development of an institutional framework within specific university departments, which was conducive to the consolidation of Communication research in Spain, then no less could be said of the intense secondary institutionalisation process embarked upon by the scientific community through the creation of scientific associations, research groups, and specialised journals and editorial collections. The history of this process largely remains to be written, however. Spanish scientific societies linked to the sphere of Communication emerged with the creation of the Association of Communication Researchers in 1982, which became defunct in the 1990s after a change of name in the previous decade to the Association of Communication Researchers of the Spanish State. The Spanish Association of Semiotics was created in 1983, the Catalan Society of Communication a year later, and the Spanish Association of Film Historians in 1988. That tendency to specialise has been a characteristic feature since then, with the emergence of associations grouping together specialists in Journalism (1990), History of Communication (1992), Educational Communication (2000), Public Relations (2004), Political Communication (2008) and Health Communication (2013), many of which are also open to the respective professional sectors. To these it is necessary to add others that are more generic and generally regional in their scope8. That secondary institutionalisation task undeniably culminated in the creation of the Spanish Association of Communication Research in 2006, and we are now beginning to see evidence of its boost to Communication research in Spain (Rodríguez Gómez, 2016). Much scarcer, not to say non-inexistent, is an analysis of Spanish Communication research centres and groups. Regarding the research centres, we cannot fail to highlight the importance that bodies such as Telefónica’s Foundation for the Development of the Social Function of Communications (Fundesco, 1969), the Government of Catalonia’s Communication Research Centre (1987) (Jones, 1998b), and the Autonomous University of Barcelona’s Communication Institute (1997) have had in this field. And, regarding the research groups, only recently has attention begun to be focused on them (Mancinas et al., 2015; Tur, 2017). Although we now at least have a repository that identifies 213 of them in Spain currently9, the task of analysing such fundamental aspects as their level of activity, scientific production, internal structure, operation, etc. has yet to be done. As far as editorial activity in the sphere of Communication is concerned, the first specialised collections by imprints such as Gustavo Gili, Paidós, Cátedra and Ariel, among others, were crucial to the take-off of Communication research in Spain from the mid-1980s, yet there are practically no studies on their development and impact on the scientific community (Giménez Toledo & Tejada, 2012). In this secondary institutionalisation process, the aspect that has been covered the best is that of specialised academic journals, a sector that has developed exponentially in recent years, coinciding with the success of the paper as a highly valued scientific Communication format for the purposes of accrediting research track records since the implementation of the Academia programme in 2008 (see section 3 of this work). The impressive status gained by scientific journals in terms of attributing prestige to researchers through a kind of metonymical transfer, where the citations received by all the works published in a certain period (the impact factor) automatically become indicative of the value and quality of each one; in our view, such an impressive status held by journals explains why the interest in analysing them has focused on the discussion and refinement of indicators (editorial protocols, database visibility, experts’ opinions, accessibility, etc.) that help to sort them into hierarchised catalogues, the well-known impact rankings (Giménez Toledo & Alcaín, 2006; Fonseca, 2011; Giménez Toledo, 2011; Baladrón & Correyero, 2012; Delgado & Repiso, 2013). The possibilities for personalised scientific dissemination that the new digital environment has generated (institutional repositories, academic social networks, etc.) may indeed be sending shockwaves through the privileged gatekeeper position that journals in the knowledge distribution system have enjoyed for so long, but the fact that it is now feasible to determine the specific impact of each researcher, with the h-index leading the way10 (De Pablos, Mateos & Túñez, 2013; Túñez, 2013; Gómez Calderón & Roses, 2015; Costa, 2017), does not detract from the fact that the researchers’ assessment and prestige is going to continue to depend on parameters of the type mentioned above. And, having lost all sense of humility, that kind of ranking fever will end up giving us lists, not of journals, but of more or less cited authors, like those lists that are already popular in other disciplines (Salgado & Páez, 2007). This may feed academic vanity – or even fuel rivalry between colleagues – but it is a somewhat vulgar criterion for fairly assessing researchers’ merit and influence. Returning to the journals, the fact is that we are neglecting to analyse researchers’ track records and contributions because we are so preoccupied with the game of rankings, visibility management or strategies to optimise the incipient practice called research marketing (Túñez, 2013, pp. 54-55). Despite being a more important task, such analysis in Spanish Communication research is scarce (García Galindo, 2013). To fill this gap, first we should stop thinking of them as scientific metrics fodder alone, and once again concern ourselves with the institutional role they play, or at least should play, in the scientific ecosystem, which is not that of serving as mere channels for knowledge dissemination, but that of spearheading epistemological debates and guiding research at any given time. And that requires the restoration of their collective intellectual status, which they should never have lost in the first place, instead of reducing them to the dullest role of simply being citation generators11.

3.4. Internationalisation of the scientific community

Of the various “ANECA effects”, the internationalisation of the scientific community is perhaps the one on which we have more rigorous evidence, at least in regard to the evolution of the presence of Spanish researchers in specialised journals published outside Spain, and especially in those included in international databases and hierarchised lists according to their impact factor. The data provided by Fernández Quijada and Masip (2013) on works signed by Spanish authors between 1980 and 2010 in international journals included in the Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (AHCI) showed a slow progression with numerous ups and downs over the 20 years from 1986 – there were no records beforehand – to 2006, in which 18 works were identified, the highest number in that period. From that moment on, the evolution was steady, though there was a clear turning point in 2009 when 60 articles were published, more than double the previous year’s number (28). Although these data refer exclusively to journals included in databases in which publications from the Anglosphere are overrepresented, they are nevertheless indicative of a change in behaviour among Spanish Communication researchers. It is therefore tempting to link that to the conditions set by the reward system implemented in 2008, which is based around six-year rewards, and to university teaching staff accreditation. The results from a work on the presence of Spanish authors in journals not published in Spain but included in the Communication catalogues of JCR and Scopus in the decade spanning 2003-2012 have provided additional evidence in the same sense (Martínez Nicolás, 2014). Thus, of the 196 contributions recorded, 29% of them were published between 2003 and 2008, and 71% of them corresponded to the period of full implementation of the Academia programme, with such a shift in progression between 2008 and 2009 – when the number of articles rose from 19 to 41 – that it seems difficult not to attribute it to the impact of those new university accreditation requirements. The presence of Spanish authors in journals not published in Spain is not the only bibliometric indicator for assessing the intensity of internationalisation of Spanish Communication research. Indeed, in that very sense, various authors have, for example, pointed to the increase in collaboration between Spanish and foreign researchers (Fernández Quijada & Masip, 2013, p. 19-20); to the presence in Spanish journals of authors working in other countries, although they are generally assigned to Latin American universities (Escribà & Cortiñas, 2013, p. 39-40); and, lastly, to widespread familiarity with international bibliography, albeit with a curious behaviour that causes it to be cited much more often (54.2% compared to 10.4% of the references) in papers appearing in journals published outside Spain than in papers appearing in journals published in Spain (Casado & Fernández Quijada, 2015, p. 80). Be that as it may, and regardless of how significant those indicators might be, a proper assessment of this internationalisation process requires us to expand the focus beyond strict scientific production and, above all, to avoid falling into the trap of assuming that it is a recent phenomenon. Indeed, the dissemination of works on international circuits (journals, conferences, etc.) and participation in teams and projects in this sphere are fundamental factors of the scientific community’s internationalisation, and the position gained over a 10-year period in this respect probably bears no parallel to previous periods. However, the international projection of Spanish Communication research has a long history that, to a large extent, has yet to be reconstructed. An essential chapter of that history would have to be the intense activity undertaken to internationally connect our scientific community when this disciplinary field was in its infancy in Spain12.

4. Characteristics of scientific production in the field of communication

Changes in the institutional context and in the structure of the academic community cannot have failed to affect the general direction of scientific production in the field of Communication over the past 25 or 30 years; that is, the practices, modes and epistemological choices applied to the task of generating knowledge in this disciplinary field. This is almost certainly the most controversial and poorly addressed aspect of the Communication research system in Spain, with the exception, that is, of some of its most objectifiable characteristics (publication formats, collaborative networks, etc.). Poorly addressed, above all, when it is about obtaining a general epistemological framework (dominant objects of study, theoretical approaches, methodological perspectives, etc.) where the scarcity of data and the divergence in the categories used in the available works make reconstruction difficult. And controversial because, based on deficient empirical evidence, attempting to point out the observed changes in trends and to highlight the factors that could have caused them are risky exercises. Nevertheless, and however imperfect the result might be, it is a risk that we should start to assume.

4.1. Success of the scientific paper

Growth in the publication of works in specialised Spanish journals – i.e., scientific papers – has been relentless since the 1980s. In this respect, the most exhaustive census may be the one done by (2016, p. 36) between 1980 and 2013, which shows two clearly marked temporal turning points: one in 1996, the year when published articles had comfortably doubled compared to 1995 (from 125 to 265); and another around 2008, when they had increased 40% compared to 2007 (rising from 514 to 715). In the 1980-1995 period, the annual mean of works in Spanish Communication journals was slightly higher than 63, compared to around 20 in the early years (1980-1984), a low figure for obvious reasons. The sharp increase recorded from 1996 meant that the mean number of papers per year rose to 440 between 1996 and 2007. And between 2008 and 2013, that figure practically doubled to a mean of 800 articles per year, and did not fall below 700 published works in journals published in Spain alone in any of the years13. It was in that latter period that we saw the unquestionable culmination of the success of the scientific paper in Spanish Communication research. When looking for an explanation for that behaviour, it is once again tempting to link the recorded change in trend around the mid-1990s to the multiplication of Communication studies and to the consequent growth of the scientific community that, especially in the new specialised centres, created journals to provide an outlet for their scientific production. And it is just as tempting to link the step change observed from 2008 to the implementation of the Academia programme for teaching staff accreditation and the value placed precisely on research disseminated via scientific journals from that time on. Additional data pointing in the same direction are those relating to the evolution in the number of Communication journals published in Spain between 1980 and 201514. To the five existing titles in 1990, new ones were slowly added throughout the 1990s, reaching a total of 29 in 2001. In 2002, coinciding with the creation of ANECA, a small jump was observed, taking the number of journals published in that year to 34 (five more than in 2001). The number then continued to rise by one or two per year until 2009, when it reached a total of 45. Eight journals were launched in 2011 and a further 10 in 2011 (i.e., 18 in two years, as many as those founded in the decade spanning 2000-2009). Then, with the Academia programme in full swing, growth in the number of journals was staggering; by 2015, 69 titles were being published in Spain: a breeding ground that was ripe for, or indeed spurred on by, the success of the scientific paper. The article is obviously not the only format that researchers use to disseminate the results of their work. It co-exists alongside other scientific products (books, conference presentations, reports, grey literature, etc.) that have indeed been relegated to subsidiary positions in the scientific publication ecosystem due to the preferential, if not dominant, position conferred on the article by the institutional scientific research reward system under which the six-year rewards and accreditation were established (Soriano, 2008 and 2017)15. However, beyond that undeniably interesting issue, thought should perhaps be directed instead at determining how the scientific paper format is effectively conditioning how Communication research is being conducted in Spain. Although the indicators probably need to be refined and supported by empirical evidence that is more robust and diverse, we do have a number of suggestive hypotheses in this respect. The issue has been raised especially vigorously by Goyanes (2015 and 2017), for whom the scientific paper and journal culture, fuelled by that reward system, encourages standardised research that leads to what he calls isomorphism; that is, “the collective thought that, by us all doing the same, we are more right” (Goyanes, 2017, p. 29). This might range from seemingly minor manifestations such as the requirement to shoehorn the scientific papers into the IMRaD structure – one of those “new grammars” of scientific writing about which Sierra (2016) suggests we should be cautious – to others that are more important such as thematic hyperspecialisation, scientific tribalism, methodological fetishism or, lastly, the bureaucratisation of research activity that tends to put the cultivation of the scientific paper on a pedestal. Indeed, the conciseness that the scientific paper requires would tend to favour such hyperspecialisation among researchers, prompting them to resolve any infinitesimal shortcomings in available knowledge – the detection of “gaps” as referred to by Goyanes (2017, pp. 57-65) – driven by a kind of hubris that fuels the formation of academic tribes devoted to learning more and more about less and less until arriving at a point where they know absolutely everything about nothing. That “barbarity of specialism” about which Ortega y Gasset warned us nearly a century ago, when he could not have imagined the magnitude that the “revolt of the [academic] masses” would reach, is now also spurred on by a publishing business, that of journals, which is ready and prepared to provide any of those tribes with their own specialised journal. This purpose, this new ecosystem, is creating a scientific community in which there is beginning to be an abundance of hedgehogs and a shortage of foxes16, and which, by indulging that methodological fetishism that Goyanes (2017, pp. 39-48) also denounces, seems to be abdicating its duty of epistemological vigilance and reflexivity that scientific work requires (Bourdieu, Chamboredon & Passeron, 1973; Vassallo de Lopes, 1990) in order to pursue poorly understood yet highly rewarded research productivity.

4.2. Take-off of collaborative research

Collaborative research practices giving rise to works signed by several authors enable us to determine the level of integration and formation of epistemic networks – generated by shared epistemological orientations: objects of study, theories, methods, etc. – in a scientific community. The reconstruction of academic networks deployed in doctoral thesis examination panels, for example, also serves that purpose (Repiso, Torres & Delgado, 2011a, 2011b and 2013; Casanueva & Caro, 2013). Very little attention has been paid to this issue in Spanish Communication meta-research. Moreover, what little attention has been paid to it focuses on the case of signed articles in journals and thus does not consider the diversity of formats that scientific collaboration among researchers may take (collective books, conference presentations, report elaboration, etc.). And although there is not much of it either, available empirical evidence on the evolution of the authorship regime of works published in scientific journals suggests a significant increase in collaborative practices in Communication research in Spain, with a clear turning point around 2008. Fernández Quijada and Masip (2013, pp. 19-20) analysed the articles signed by Spanish researchers in 43 specialised journals between 1980 and 2013, and, in the case of journals published in Spain, they found that single-author works accounted for approximately 85% of all those published between 1980 and 2007, a trend that changed very clearly from 2008 with a sustained increase in the number of multi-author papers in the three years that followed. Between 2007 and 2008, collaborative research rose from 17% to 22%, and, from that moment on, it continued to increase, reaching nearly 30% in 2010. In other words, that way of working practically doubled in a space of four years (2007 to 2010), breaking a pattern that had hardly changed over the previous 25 years. The pattern for publishing in international journals was mostly that of multi-authorship across the period analysed by those authors, although the volume of works makes the behaviour observed between 2007 and 2010 especially significant because there was a constant increase in the proportion of papers signed by various authors over those four years, which went up from 60.9% to 67.4%. A similar result was obtained from a recent study on five journals published in Spain over the 1990-2014 period (Martínez Nicolás, Carrasco & Martínez Fernández, 2017). The study showed that the growth of co-author works was slow yet sustained between 1990 and 2009, rising from 10% in the 1990-1994 period to just over 25% in the 2005-2009 period. However, in the last five-year period (2010-2014), they accounted for more than half of all the works. Although referring exclusively to Advertising research, Baladrón, Machado and Correyero (2017, p. 6) provided additional data on the then recent boost to collaborative practices that, in this specific sphere, had shot up between 2011 and 2015, also accounting for around 50% of works published in each year of that period. Continuity in the progression of collaborative research published in scientific journals since 2008-2009 suggests that what we have before us is a consolidated trend, from which it is possible to infer a structural change in a scientific community that, in recent years, has become more integrated and has formed more academic networks. Here too it is tempting to link the rise in such collaboration practices to the requirements established for accreditation and for the six-year rewards for research, with the latter being essential merits for promotion at Spanish universities, but not for access to them. The temptation to establish that link has likewise given rise to a kind of cynical hypothesis according to which the substantial growth in multi-author works is not so much a reflection of the formation of academic networks as it is of spurious authorship exchange. Indeed, being an object of value so sought after by the institutional reward system (six-year rewards, accreditation), it would surely be a waste for the benefit that publishing in scientific journals might bring to be attributed to a single author, and especially so under circumstances of tough competition in terms of getting a paper published in such journals. Consequently, and according to this cynical hypothesis, that institutionally induced picaresque approach may have enhanced the vigour displayed by co-authorship in Spanish Communication research in recent years. Such authorship exchange (“I’ll credit you, you credit me”) may of course occur. But before resolving the issue so expeditiously, it is worth taking note of other data that may help to assess it more appropriately. With records of 14 Spanish journals for the years 2007 and 2008, Fernández Quijada (2010, p. 564) found that “most co-authorships pertained to the same institution, and inter-institutional collaboration was an infrequent phenomenon [...]”. However, more recent data indicate that the pattern has clearly been changing since then. Despite the limited nature of our sample, our analysis of the authorship regime in five Spanish journals between 1990 and 2014 (Martínez Nicolás, Carrasco & Martínez Fernández, 2017) showed a noticeable change in trend over that 15-year period – the results were not significant for the 1990s – in which works signed by researchers assigned to different universities (inter-institutional) went up from 6% in the 2000-2004 period to nearly 23% in the 2005-2009, and to more than 34% in 2010-2014. In other words, in the latter years covered by that study, a third of collaborative research published in the analysed journals was inter-institutional, whereas just a decade earlier, it did not even reach a tenth. Furthermore, collaborative works generated by funded projects in that same 15-year period slowly increased from 8.6% in 2000-2004 to 11.5% in 2005-2009, and rose sharply to 27.6% in 2010-2014, nearly three times higher than in the previous five-year period. This coincided precisely with the recognition of Communication as a specific field in the State RD&I Plan, and with the growth in the number of funded projects in the area. Flying in the face of that cynical hypothesis of authorship exchange, these data showed, for at least the last decade of the period covered, that the scientific community committed to the sphere of Communication was undergoing a rigorous process of integration, firstly with the formation of multi-centre networks that were probably less permeable to such a spurious exchange than in-house networks formed within departments and faculties, and above all with the organisation of research teams that, as a general rule, were also multi-centre ones. The latter were of course necessary in order to be able to plan and execute projects of the appropriate magnitude to successfully compete in the most financially and curricularly beneficial calls for applications: those of Spain’s State Plan and the various European programmes. In sum, these factors should be taken into consideration if we want to fairly assess the rise in collaborative research over the past 10 years.

4.3. General direction of communication research

Content analysis, as it were, of scientific production in certain disciplinary fields can take two directions, neither of which has been addressed fully or in depth by Spanish Communication meta-research. One of those directions – demanding and generally applied to domains or specific problems – aims to assess the contributions to, and the weaknesses and pending challenges of, conducted research, taking it as the starting point from which to propose heuristic alternatives of a theoretical or methodological nature to mitigate those weaknesses and address those challenges. In reference to Spanish Communication research, works of this type are few and far between. Indeed, we only have a small number, and these are on certain theoretical perspectives (e.g., Vicente & López Rabadán, 2009; Ardèvol, 2015; and Varela, 2016, on framing studies) or specific spheres (Martínez Nicolás, 2006, on Journalism; Almiron & Reig, 2007, on the political economy of Communication; De Miguel, 2007, and García Jiménez, 2007, on Sociology and Communication theories, respectively; Martínez Nicolás, 2007, on Communication policy; and Fernández Viso, 2012, and Marí, 2017, on Communication for social change). The other direction – no less demanding in its execution – reports on the general state of play of research by attending to the diverse epistemological components it comprises: knowledge interests or objects of study that attract researchers’ efforts; the theoretical approaches taken to address them; the research modalities – theoretical, empirical – they employ; and, lastly, the methods and techniques to which they resort. Here too there is a shortage of works, since we researchers are comfortably set on uncovering those aspects of scientific production that bibliometricly stare us in the face (authorship patterns, universities of provenance, researcher networks, citation styles, use of references, etc.) instead of shedding light on other aspects that require the interpretative gaze of an expert in the field (objects of study, types of research, methodological quality, etc.). Although they focus on the analysis of production published in scientific journals, here too we have a number of valuable works (Castillo & Carretón, 2010; López Rabadán & Vicente, 2011; Martínez Nicolás & Saperas, 2011; Martínez Nicolás, 2014; Baladrón, Correyero & Manchado, 2014; Martínez Nicolás & Saperas, 2016; Martínez Nicolás, Saperas & Carrasco, 2016 and 2017; Piñeiro, 2016; Barranquero & Limón, 2017; Caffarel, Ortega & Gaitán, 2017), but the disparity between the categories used to account for those epistemological components hinders the task of reconstructing the characteristics of Spanish Communication research over the past 25 or 30 years. Despite that, we shall endeavour to present those general directions.

4.3.1. Diversification of knowledge interests

The internal diversity acquired by the scientific community as a result of the establishment of distinct Communication degrees in the early 1990s (see section 2.2.) boosted research in those Communication and Media spheres, such as Advertising, Public Relations and Corporate Communication, and Audiovisual Communication, which had traditionally been relegated to a subsidiary position compared to the attention paid to the field of Journalism. The data provided by Castillo and Carretón (2010, p. 311) for 2008 pointed to the fact that Journalism was still the predominant “thematic area” in 10 Spanish journals (22.3% of the papers), way ahead of those dedicated to Television (13.7%), Advertising (9%) and Public Relations (7.4%). However, the results obtained from the analysis of a sample of 1,000 articles signed by Spanish researchers in five specialised journals published in Spain (Martínez Nicolás, Saperas & Carrasco, 2016 and 2017) showed a gradual diversification of those knowledge interests over the past 25 years. Of the works recorded in the 1990-1994 period (in the only two active journals of the five selected), 70% studied issues relating to Journalism, though the proportion began to fall continuously from the mid-2000s (2005-2009 period) to just above 40% in the final period (2010-2014), when articles on Audiovisual Communication accounted for a quarter of all published articles, and those referring to Advertising and Public Relations accounted for 20% of that whole. A similar trend was observed in works by Spanish authors in international journals included in JCR and Scopus between 2003 and 2012 (Martínez Nicolás, 2014). Around 25% of those published in 2012 addressed Journalism, but more than 42% studied aspects of Audiovisual Communication, Advertising or Public Relations, the latter of which was particularly prominent (20% of the international papers in that year). Over the past decade, Spanish research has paid special attention to new phenomena spawned by the advent of the Internet and the digitalisation process. Of the articles in Spanish journals in 2008 recorded by Castillo and Carretón (2010, p. 311), 8.2% of them addressed those topics. However, in the 2008-2014 period, they accounted for nearly 20% (Martínez Nicolás & Saperas, 2016, p. 1373), a percentage similar to the one reached in 2012 in international journals (Martínez Nicolás, 2014), where works by Spanish authors focusing on that digital environment accounted for nearly 25% of those published. In sum, everything seems to point to a gradual diversification of the knowledge interests of a scientific community that no longer primarily or predominantly puts its efforts into journalistic research.

4.3.2. Centrality of media content analysis

Beyond the Communication and Media spheres (Journalism, Advertising, Public Relations and Corporate Communication, Audiovisual Communication, and the Internet) addressed by researchers, a proper consideration of knowledge interests requires, at the very least, an observation of the elements of the Communication process – or of the Communication system, if you will – that are specific objects of study: professionals, businesses and markets, content, technologies, audiences and effects, or other aspects of a reflexive nature such as teaching or Communication research itself. Regrettably, these analysis categories, which are crucial to being able to answer the basic question of what is being investigated in a certain discipline (Martínez Nicolás & Saperas, 2011, p. 108), do not seem to have made their mark on Communication meta-research, so very little can be said about them without the support of enough empirical evidence. Be that as it may, the little research that is available (Martínez Nicolás, Saperas & Carrasco, 2016) shows that media content analysis studies occupy a central position. Generally speaking, such studies usually take the shape of journalistic discourse analysis (news coverage or news treatment of population groups such as youths, immigrants and marginalised groups; situations and events such as health crises, conflicts and elections; spheres such as politics, economy and justice; and phenomena such as gender-based violence and climate change), though some studies do examine other media products (advertisements, television series, film, documents on digital forums, etc.). The data obtained by Zugasti (2013, p. 69) from seven Spanish journals between 2002 and 2012 corroborated that same trend in the specific field of the history of Communication, in which nearly 40% of all the papers published were studies focusing on the “message”. But what is perhaps most significant in this respect is that the primacy of Spanish researchers’ interest in analysing content continued to grow over time, causing a relative decline in attention paid to other elements of the Communication process (Martínez Nicolás, Saperas & Carrasco, 2016). Thus, in the early 1990s (1990-1994 period), four out of 10 articles in the Spanish journals analysed were about media content, but five of them focused on aspects relating to business (organisations’ operation, market structure, sectoral public policies, etc.) and professional spheres (production processes, ethical and deontological issues, labour profiles, corporate associations, etc.). From the early 2000s, there was an accelerated shift in the recorded objects and, in the most recent period (2010-2014), six in 10 of the papers published were on content, while only two in 10 were on Communication businesses and professions, standing at the same level as studies on audiences and on reception processes, which rose in the last period from an almost token presence in the two previous decades. Consequently, the data pointed to research that, focusing basically on generating knowledge about the discursive dimension of Communication, overlooked the sphere’s institutional (businesses, markets, professions, public policies, etc.) and social dimensions (audiences, reception, effects and influence of Communication), and shifted towards content. And that trend only seemed to pick up pace in the mid-2000 (see Martínez Nicolás, Saperas & Carrasco, 2017, p. 159, specifically for journalistic research).

4.3.3. Primacy of quantitative empirical research

A similar decline was observed in theoretical research (aimed the discussion of concepts, theories and methods, or at the presentation of the state of play of certain objects or domains) over the past 25 years, during which the empirical direction of Spanish Communication research has been considerably strengthened (Martínez Nicolás, Saperas & Carrasco, 2016 and 2017 for the case of Journalism). According to the results obtained from those works, over the 15 years between 1990 and 2005, articles addressing theoretical problems accounted for around 30% of those published, but they had practically disappeared from the five Spanish journals analysed in the last decade, when they only accounted for one in 10. Methodologically, a strong trend was observed in those works to adopt standard Social Sciences techniques, be they quantitative or qualitative, to the detriment of research based on an analysis of documentary resources (historical archives, corpora of legislative or legal doctrine, codes of ethics, policy or business reports, etc.), which, until the early 2000s, had been the predominant procedure for generating empirical data in Spanish Communication research. From then on, works resorting to those standard techniques (content or discourse analysis, surveys, in-depth interviews, focus groups, etc.) predominated, with a clear preference for quantitative research, which accounted for just over half of the empirical articles published in the 2010-2014 period, a much higher proportion than the 20% that employed qualitative techniques. Practically identical results were obtained by Castillo and Carretón (2010, p. 313) for 2008, when 54% of the selected sample were quantitative studies and 20% were qualitative ones. In contrast, Zugasti (2013, p. 71) showed that, in studies on the history of Communication, “the qualitative approach overwhelmingly outweighed the quantitative one” (nearly 86% of the studies included in the seven Spanish journals between 2002 and 2012). However, as we noted at the time (Martínez Nicolás & Saperas, 2011, p. 119-120), that may have been due to the propensity to consider all research that does not explicitly resort to quantitative empirical techniques as qualitative research. Indeed, works like these cannot be relegated to a kind of methodological orphanhood (where they are neither quantitative nor qualitative). And that is why our proposal is to classify them as research based on documentary sources, provided such sources are duly identified and rigorously exploited. Be that as it may, quantitative approach appears to have become firmly embedded in Spanish Communication since the mid-2000, years, with a sharp rise in the past decade.

4.3.4. Marked improvement in methodological quality

In one of the few studies specifically evaluating “methodological approaches” in Spanish Communication research, López Rabadán and Vicente (2011) analysed the methods used in articles published between 2000 and 2009 in four specialised Spanish journals, and they found that, in 43.8% of them, “it was not possible to locate any type of reference to how the research was conducted”. They also found that only 23.8% of them included “a detailed chapter on methodology” and that 32.5% of them “incorporated the odd reference to these aspects” (López Rabadán & Vicente: 2011, p. 9). Albeit with a much narrower timeframe referring only to 2008, Castillo and Carretón (2010, p. 313) observed that “it is surprising to find that, in 23.9% of the articles, there was no method at all”. There is, therefore, more or less robust evidence of the fact that empirical Communication research in Spain was severely lacking in terms of methodology, and that is something we also found in an analysis of a 1998-2007 sample of Spanish journals (Martínez Nicolás & Saperas, 2011, p. 19). However, if we focus attention on the evolution of this parameter, there are reasons to argue that this anomalous situation began to be rectified from the mid-2000s (Martínez Nicolás, Saperas & Carrasco, 2016). The data obtained from articles published by Spanish researchers in five journals published in Spain, grouped into five-year periods, showed that, in the 1990s, the relationship between works that were methodologically deficient and those that showed signs of being methodologically sound remained unchanged, with the proportions standing at 45% and 55%, respectively. In the 2000-2004 period, a slight change in trend was observed in this respect, which became fully consolidated in the last decade. Thus, the proportion of articles with basic methodological shortcomings fell to 27.9% in the 2005-2009 period, and fell again to 13.9% in the most recent period (2010-2014). It is, as we said, a sign of marked methodological improvement. What was exclusively considered when assessing the empirical works was whether or not they resorted to some standardised technique (content analysis, surveys, focus groups, direct observation, etc.), or followed suitable protocols for the analysis of documentary sources. In the case of the former, no assessment was made of whether or not those techniques were correctly applied in accordance with the procedures established by the methodological literature. Be that as it may, and albeit only a sign, the available data continue to evidence the industrious efforts that Spanish Communication researchers have been making to adapt their works to increasingly demanding methodological standards.

5. Final thoughts

A review of what has happened in Spanish Communication research over the past 30 years provides us with a picture of a period in which the institutional framework supporting scientific activity in this field has changed profoundly, with two clear turning points: one at the beginning of that period when the University Reform Law was passed (1983), which set in motion a veritable process of university modernisation in Spain; and another in the mid-2000s, after the implementation (2002) and generalisation (2008) of the accreditation system for access to and promotion in university teaching staff faculties, which was more demanding in terms of research requirements. These general institutional changes, accompanied by others that specifically affected the sphere of Communication (multiplication of university study programme offerings, establishment of distinct degrees, recognition of Communication in the State RD&I Plan), would leave a clear mark on the structure and characteristics of the academic community and on the general direction taken in Communication-related scientific production in Spain. Indeed, the impact was so strong that, in light of the data now available, the proposed periodisation of the evolution of Spanish Communication research (Martínez Nicolás, 2008) should probably be revised to situate, around the mid-1980s (after the University Reform Law had been passed), the beginning of a phase that, albeit with the inertia inherited from the time the discipline was founded in Spain, represented a profound renewal (expansion of the secondary institutionalisation of the scientific community, connection with international traditions and lines of work, etc.) that would continue for 20 years up to the mid-2000s. Thereafter, the empirical evidence that we have gathered in the latter years of intense meta-research points to a change of cycle. Given this state of affairs, the key issue is to figure out what path Communication Research in Spain has followed since then. That change of cycle has brought us a more diversified scientific community, with the gradual yet stable incorporation of women into research activity in this sphere. And, in view of the rate of growth in the number of doctoral theses being produced, that community now has sufficient critical mass to further develop the field, even though the opportunities for professionalisation within it have dwindled due to the fatefully combined impact of financial cutbacks applied to universities and the greater stringency of access requirements. And, as the backdrop to all of this, we have the unwavering demands for excellence, quality and innovation that are characteristic of the neoliberal managerialism to which managers of university policy in Spain seem so attached (Girotto, Mundet & Llinàs, 2013; Quirós, 2016). On a strictly epistemic level, that change of cycle seems to point to a type of research that is more diverse in terms of its interests after the long-standing predominance of the journalistic tradition in Spain. It also suggests that such research is well disposed to generating knowledge on the changes brought about by the advent of the Internet and digitalisation in all areas of life in contemporary society, an interest that undeniably puts us in a privileged position in the general field of Humanities and Social Sciences. And, in an especially valuable way, that change also points to an improvement in the rigorousness of the research we are doing, albeit with certain aspects that need to be rectified (the bias towards content, the near methodological monoculture of quantitative content analysis, minimal recourse to techniques to obtain data from living sources, such as the use of surveys, experimental designs, focus groups, direct observation, etc.). Furthermore, our research has proven capable of making a place for itself in the international ecosystem of knowledge dissemination (journals, conferences, projects, etc.). Yet all this still comes at a cost of yielding to, or perhaps only converging with, an internationally imposed scientific standard (Carrasco & Saperas, 2014 and 2016). Although we have only been immersed in this state of affairs for no more than 10 years, it may actually be announcing the advent of a new scientific culture in Spanish Communication research, the take-off of which will especially involve committed generations of younger people, driven to make research a way of living, a kind of necessary routine in an environment bedevilled by the competitive hypertrophy that drives the established reward system. However, as we internalise the fact that researchers have no option but to increase productivity, seek impact or climb in the rankings, we will shape a scientific community in which good, technically proficient professionals will abound. That said, it will probably be a scientific community in which we will end up pining for a more ambitious intellectual attitude, that is, one that makes Communication research an activity committed to criticism and social intervention rather than a vacuous exercise aimed at accumulating increasingly inane knowledge.

Having embraced, adopted and adapted from many longestablished research disciplines including sociology, psychology and education, health professions education (HPE) research has now become a research field in its own right. This trajectory has transported HPE beyond its initial narrow positivist frame and supremacy of the biomedical model to take up increasingly more critical qualitative approaches to HPE inquiry. However, while qualitative methodologies such as discourse analysis, constructivist grounded theory and phenomenology have flourished, institutional ethnography has to date garnered little attention in this area of research. Institutional ethnography is a critical theory/methodology, with a particular focus on people’s everyday lives and how their lives are organized and coordinated by institutional forces. Use of institutional ethnography has prospered in clinical healthcare research, particularly in nursing, as well as in social work and education, bringing about useful insights and tangible change for frontline workers [1, 2]. We argue that institutional ethnography is a unique approach to inquiry that is especially suited to research in HPE. In this article, we open a new dialogue in HPE by offering an opportunity to delve into institutional ethnography’s unique conceptual underpinnings. We do not intend this article to be an institutional ethnography ‘how to’ per se; for this, readers are directed to Ng et al. [3], Campbell and Gregor [4], Smith [5] and Smith [6]. Rather, we aim here to make this innovative approach to inquiry more accessible in HPE, by laying the foundations to enable readers to consider it and to further the use of institutional ethnography in the field of HPE research. As we considered this manuscript, we could see multiple, current research opportunities for institutional ethnography in modern HPE. This discussion of institutional ethnography and its potentiality in relation to HPE is timely, especially given the context of contemporary HPE in industrialized and industrializing societies. While the HPE field is regulated differently depending on the national and political context, in addition to the discipline-specific field (which is often in isolation from other disciplines), the ties that bind HPE is where the ethos of neoliberalism and ‘new public management’ are visibly involved in reorganizing HPE [7].1 Neoliberalism refers to the ideology that the ‘market,’ and hence market-based solutions, is the most efficient and effective way to address public sector problems, whereby new public management is the method by which the ideology of neoliberalism is put into practice. As Griffith and Smith explain, ‘applying what has come to be called new public management has involved the adoption and adaptation of strategies and textual technologies that revolutionized corporate management during the 1980s and 1990s’ [7]. The creep of neoliberalism and new public management, for example, is evidenced by the imposition of evidencebased guidelines in the healthcare arena broadly and a focus on achieving and documenting ‘competency’ in HPE specifically. These shifts have brought new challenges to HPE, particularly in the context of powerful prevailing discourses that result in pervasive standardization throughout HPE [10] (e. g. in summative assessments) with unintended consequences such as the development of ‘tick box’ style questioning by students and the fear that such a digitized form of questioning may translate into their interactions with real patients. Exemplifying the increased presence of these discourses into the world of HPE, traditional management terms such as accountability and efficiency have become quiet murmurs or even common parlance for those involved with this work. It is with these shifts in HPE in mind that our discussion of institutional ethnography follows. In the first section, we provide an in-depth overview of institutional ethnography, which includes key methodological and theoretical tenets of this ‘alternative’ sociology. In the second section, we provide three short analytical examples in order to provide 1 These ideologies have become pervasive, or as some have noted, a ‘hegemonic hybrid’ [8, 9], throughout society. As Clarke and Newman explain [9], most ‘Civil Service’ arenas are organized by market- oriented relations of coordination and control in an attempt to achieve ‘the ‘lean state”. readers with a sense of institutional ethnography in practice. This is followed by the third section whereby we discuss some potential areas where institutional ethnography can be used to explore and explicate contemporary HPE practices. We end this article with a brief conclusion.

Institutional ethnography

For a thorough understanding of the complexities of institutional ethnography, it is necessary to know something about Dorothy Smith, contemporary sociologist and founder of this approach to inquiry. Institutional ethnography was borne out of her life’s work critiquing mainstream sociology, rejecting what she considered its inability to start in the real world or to explain the ‘bifurcation of consciousness’ that she experienced in her conflicting worlds of academia and parenthood [11]. She turned in particular to the teachings of Marx’s materialism and Garfinkel’s ethnomethodology, but also to Foucault, Mead, Bakhtin and Volosinov, and their insights on language, power and knowledge. Through her involvement in the feminist movement, she was introduced to the concept of ‘consciousness raising.’ This became key to her own sociological insights, encouraging people ‘to speak from themselves and their experience’ [11]. These early influences and experiences culminated in the development of institutional ethnography. Institutional ethnography is a critical qualitative theory/methodology that ‘starts from people’s everyday local experience and explores the translocal that is present in and organizes their everyday’ [12]. In institutional ethnography, we are interested in ‘how things work’ and ‘how they are actually put together’ as opposed to ‘what happens’ or ‘why things happen’ [11]. The emphasis is on what people do—their work broadly conceived—and what individuals say and know about their work as expert knowers and doers. This focus on work contrasts with more traditional forms of qualitative inquiry that tend to be organized by positivist tenets aimed at providing what DeVault and McCoy refer to as a ‘window on the informants’ inner experience’ (cited in [5]); it is this expert knowledge of work that provides the entry point into the inquiry [11]. By starting from this distinct position where institutional ethnography does, in people’s everyday worlds, it is also fundamental to point out that it does not, in contrast to some more conventional sociological and other research practices, start or end in theory.2 Starting in theory and using theory in traditional ways can result in what Smith calls the ‘14th floor effect’ [12] whereby theoretical concepts stand in for the social relations that exemplify the theory. Smith writes that in more mainstream approaches to social scientific inquiry, what actually happens on the ground—what people are doing—is objectified ‘above’ local happenings based on the theoretical frame deployed, displacing the presence of people as knowing subjects and their everyday doings [6]. In order to move beyond mainstream research that tends to begin with and end in theory— a remnant of positivist ways of thinking—institutional ethnographers begin with a ‘problematic.’ A researcher’s problematic ‘sets out a project of research and discovery that organizes the direction of investigation from the standpoint of those whose experience is its starting point’ [5]. A problematic is used in institutional ethnography to direct attention to ‘a possible set of questions,’ tensions, or puzzles that are ‘latent’ in, yet arise from, people’s everyday actualities [11]. Having begun in the everyday actualities, often through in-depth interviews, participant observations, and sometimes through a researcher’s own reflection of their everyday life, institutional ethnography then moves, to ‘investigate how their activities are coordinated. It aims to go beyond what people know, to find out what they are doing is connected with others’ doings in ways they cannot see’ [5]. Institutional ethnography orients to exploring and explicating the social relations that organize that experience in the institutional setting or settings in which they exist. It is thought that individuals participate in these sites of interface often without knowing and in a way that may not be initially obvious from their own standpoint within the institutional complex where they are situated. It is important to point out that institutions, rather than referring to buildings or organizations as such, are defined by Smith as ‘complexes embedded in the ruling relations that are organized around a distinctive function, such as education, healthcare, and so on’ [5]. ‘Keeping the institutional in view,’ as per McCoy, (cited in [6]) is not only fundamental but obligatory in institutional ethnography; indeed, it is thought that a common error by the inexperienced institutional ethnography researcher when collecting and analyzing data is for their focus to remain only on what is happening on the ground. In order to keep the institution in view, a key defining feature of institutional ethnography is the mediating role that texts play in this coordination of peoples’ work; texts are viewed as being at the juncture between the everyday work people do and how everyday doings are organized and coordinated [4]. Whilst texts play a key role in many forms of critical qualitative research, institutional ethnography’s approach to texts is somewhat distinct. To provide some context, texts in institutional ethnography are viewed as ‘... definite forms of words, numbers or images that exist in a materially replicable form ...reproduces them across time and space and among people variously situated’ [13]. In institutional ethnography, texts are never looked at in abstraction, devoid of the context in which people use them but once read or used in some way, they and the discourses embedded within them are viewed as being ‘activated.’ At this point they become active ‘constituents of social relations’ ([14], for further discussion, see [5, 15, 16]) and their ability to coordinate becomes visible. The analytical intent is to explore how lives are ‘put together’ across a multiplicity of different sites [17]. Smith’s work resulted in her bringing about an entirely alternative sociology, the social organization of knowledge. She feels that established sociology looks at the lives and social relations of people as if from the outside, forefronting objectivity by ignoring what people do in their everyday lives and their subjective experience of it [11]. Smith’s social organization of knowledge offers an alternative and critical way to look at individuals’ everyday lives in context; as a sociology for people as opposed to a sociology of people. Whilst the inquiry starts in the everyday world, it is imperative, in time, to establish how this world is organized socially and ‘put together’ by the external and text mediated ruling regime or what Smith calls ‘ruling relations’. Smith [16] defines these as a ‘complex of objectified social relations that organize and regulate our lives in contemporary society.’ It is this desire in institutional ethnography to explore and explicate ruling relations and how they organize and coordinate people’s everyday lives that the focus on texts becomes particularly important; texts have ‘actual presences in people’s activities and in how activities are coordinated both as local sequences of action [between people] and institutionally.’ [5]3. In order to realize institutional ethnography as a sociology in itself and an approach to inquiry, a challenge is brought to the usual paradigm of traditional sociology and more mainstream ways of thinking. It requires both an ontological and an epistemological shift by the researcher. Ontology, according to Crotty ‘is concerned with what is, with the nature of existence, with the structure of reality as such’ [22]. The ontological shift that is required with institutional ethnography rejects speculative abstract theoretical explanations and moves instead towards what George Smith has described as a ‘sensuous world of people’s actual practices and activities’ [23]. Hence, in the social organization of knowledge, the inquiry begins in an embodied standpoint in the social, rather than beginning in abject theory [24] and seeks to trace or ‘map’ how peoples’ practices and activities are organized and coordinated by text-mediated modes of governance. Connected to this discussion of ontology is the epistemology that institutional ethnography proffers. Epistemology ‘is a way of understanding and explaining how we know what we know’ [22]. The epistemological shift in institutional ethnography involves rejecting ‘objective accounts (the view-from-nowhere type) and instead practices a reflexive way of knowing the world she or he inhabits’ [23]. Institutional ethnography strives for a way of knowing that is experiential, from the inside, rather than the truth or objective or ideological. As George Smith summarizes ‘Objective knowledge is no longer ‘the truth’ [24] and he discusses how whilst the ontological shift came about from Dorothy Smith’s reading of Marx, this epistemological shift is due to the influence of her experiences through the feminist movement; hence its common description as a ‘Marxist feminist’ approach to sociology and social scientific inquiry. Taken together, institutional ethnography’s ontological and epistemological shifts positions institutional ethnography as a non-positivist approach in that it rejects the assumption that there is a knowable (real) reality and it rejects ‘causal logic’ [25] that often results in decontextualized simplistic analyses devoid of the authorial presence of the researcher.

Institutional ethnography in practice

In the previous section, we provided readers with an overview of institutional ethnography in order to give a sense of its theoretical underpinnings and, more broadly, a sense of institutional ethnography as a project of discovery. With that said, it is important to note that institutional ethnography is still evolving, as scholars from numerous and distinct disciplines throughout the world continue to take it up, often in ongoing collaboration with Smith herself. Indeed, the decidedly political and activist roots of institutional ethnography are reflected in the social issues where it is often taken up, with the stated aim to promote social justice. This section aims to build on the previous section by providing three examples of institutional ethnography in practice. We provide short analytical accounts of three studies with a particular focus on frontline healthcare, the health work of patients and education. We do so in order to provide readers with strong exemplars of institutional ethnography and to allow readers themselves to gain a sense of what institutional ethnography has to offer in the context of HPE. In Managing to Nurse, Rankin and Campbell [1] provide an inside look at healthcare reform and restructuring practices in Canada from the standpoint of nurses who work on the frontline. They do so by ethnographically exploring what nurses do—their work process broadly conceived— and how their work is organized by what they call technologies of management and governance. Such technologies, they argue, have embedded within them neoliberal and new managerial logics that aim to make healthcare, and specifically healthcare workers, more efficient, effective, and accountable. For example, they explore tracking systems of admissions, discharges, and transfers of patients that aim to provide more effective ways of utilizing hospital resources and managing hospital costs, and clinical pathways and related technologies that aim to standardize clinical practices and determine ‘appropriate’ levels of care required based on normalized understandings of patients and needs. By ethnographically exploring the complex work of nurses and how their work is being organized and reorganized by these technologies of management and governance, Rankin and Campbell problematize ‘the apparent routineness of nurses’ work,’ work that is essential for the smooth functioning of contemporary healthcare systems, and show how such technologies have unforeseen consequences, what they refer to as ‘hidden dangers,’ that impact both frontline nurses and those they care for. Lastly, they explore and explicate how knowledge in healthcare is organized by text-mediated knowledge production processes that generate ‘official’ representation of what counts in institutional terms that while useful for the institutional complex in terms of administrative and managerial purposes is nevertheless devoid of what actually happens on the frontline [1]. More recently, Nichols et al. [26] used institutional ethnography to explore the health work—the ‘wide range of’ activities—that parents do to support the health and well-being of their child(ren) and how this family health work is organized by the institutional settings in which their work occurs. Drawing on qualitative data collected primarily through focus groups with parents from a diverse range of socioeconomic backgrounds, Nichols and colleagues shed light on how the health work of parents is mediated by a broad range of social determinants, including ‘individual, social, cultural and structural factors.’ They also show how such health work is often at ‘odds with managerial regimes’ and the biomedical model of health that discursively organizes hospital settings, despite, and perhaps in contradiction to, the pervasive focus on patient and family centred care in Canada (and elsewhere). In doing so, they provide a clearer understanding of how ‘health information, health education and healthcare interventions’ can be better designed to reflect the actual needs and lived experiences of a diversity of family types, therefore informing a critical public health. In the context of education, Griffith and Andre-Bechely, [27] begin in the everyday world of parents and trace how ‘new institutional technologies of standardization and accountability’ are reorganizing educational settings. They focus specifically on the ‘kitchen table work’ parents do as they are drawn into the many different standardizing technologies—technologies that draw parents into ‘local, state, and national political and economic priorities’ in an attempt to help their children successfully meet different institutionally mandated academic standards. While this textmediated work is changing the relationship between families and schools, Griffith and Andre-Bechely point out that in order to be successful in this changing relationship, many resources are needed. They conclude that despite the good intentions of such policies, many educational changes are a source of inequality, privileging some families over others, specifically those ‘that are able to release the mother’s time to support her children’s education, that are Englishspeaking, and that are more educated’ [27]. These brief summaries aim to provide readers some practical examples of institutional ethnography research, honouring the tradition in institutional ethnography where much of the learning around this approach to inquiry is achieved by reading and considering existing research. These studies exemplify how institutional ethnographers use theory differently than mainstream approaches to social scientific inquiry; rather than displacing and subsuming what people do in practice, theory in institutional ethnography is used to orient the ethnographer to focus on what people actually do, as they know it, live it, and experience it, but always with an analytical lens towards explicating how what they do is organized by text-mediated and regulated social organization (Smith’s ruling relations). The goal is to penetrate ‘sequentially deeper in the institutional relations in which people’s everyday lives are embedded’ [5]. As such, the ethnographies described above shed light on how the social world is put together and bring ‘into view the interface between individual lives and some set of institutional relations’ (McCoy, cited in [6]).

Why institutional ethnography in HPE and where to go fromhere?

In the increasingly crowded arena of qualitative HPE research, what does institutional ethnography offer that is not already addressed by the current methods/theories open to researchers? Why would someone choose to learn this critical approach to inquiry? One of the things that sets institutional ethnography apart from other critical modes of inquiry stems from its ability to discuss explicitly what the situation is on the frontline and its outside organizing forces, allowing discovery of what may not be questioned or even apparent through other research modalities. The health professional world, whether in training or in practice, is rife with institutional hierarchies and regulations, making the field ripe for institutional ethnographic investigations. In this light, an area that could benefit from an institutional ethnography lens is a focus on the social organization of work, including taken-for-granted work, and work settings in HPE. It is through this exploration of both visible (institutionally recognizable) and unseen dimensions of work, and how individuals are socially organized, that institutional ethnography enables genuine and meaningful social change. Through empowering individuals to recognize their position with regards to the ruling relations, institutional ethnography then allows them to challenge these positions and to consider different approaches; institutional settings can be reformed in ways that reflect the actual work processes of those on the frontline. Ng et al. propose a further benefit of institutional ethnography; they advocate that as a research modality, it offers the opportunity to combine research into the education of health professionals with research into the practice of these health professionals, championing their undoubtable interlinkage instead of separating them [3]. Both this potential for transformation and this connection of the education/practice divide is likely to appeal to the pragmatic side of health professionals, both as educators and clinicians. The current social, cultural and political landscape in health professions education provides multiple opportunities for exploration using institutional ethnography. The advent of new public management has resulted in a change in focus for what ‘counts’ for an organization, advocating for measurable outcomes and forefronting values such as accountability and efficiency. As previously stated, many studies using institutional ethnography have taken place in healthcare settings exploring how principles of new public management [7] have entered into and reorganized the work of frontline workers, including paramedics [2], nurses [1], care workers in long-term care facilities [28] and occupational therapists [29]. Webster and colleagues looked at the effect of a new policy that mandated a reduction in waiting times for patients in emergency departments, finding that the clinicians working there ‘perceived that efficiency was more important than education and was in fact the new definition of ‘good’ patient care’ [30]. This work is cited as an example where by a change in policy (activated through a text) organizes and regularizes practice [3] but with ‘hidden dangers’ [1]. The consequences of this appeared to be the emphasis on speedy rather than compassionate care where patients were perceived as obstructing this efficiency [3]. The current reliance on text-mediated forms and frameworks depicting the ‘competencies’ that health professionals should demonstrate would be an interesting space for HPE researchers. The authors of this article are particu larly interested in looking at the dominance of standardization and accountability in the assessment processes used to deem students to be ‘competent’ in their chosen profession [31]. This interest emerged from our conflicted experiences as examiners who are also clinicians, practising daily with patients whose clinical conditions are anything but standardized. Connected to this emerging trend of the dominance of managerial ‘buzz words’ such as accountability and efficiency has been the broader shift in healthcare towards practices based on evidence-based medicine (EBM), a key occupational competency taught throughout the health professions. EBM is increasingly being problematized by critical health scholars; while EBM is geared towards increasing quality of care and patient outcomes, such shifts in healthcare are also connected to the efficient use of resources in healthcare settings. For example, Rankin and Campbell [1] explain that the ‘use of a clinical pathway [a component of EBM] advances hospital efficiency by standardizing and streamlining the treatment regimens for patients with certain diagnoses.’ As healthcare and HPE become increasingly weighed down by policy and guidelines, the gap between such ‘evidence- based medicine’ and other forms of organizing and knowing healthcare and what actually happens in practice for health workers, health professions students and patients must be explored and explicated. This is especially important in this area because ways of knowing in healthcare are often characterized as ‘lines of fault’ between what is known institutionally and ‘systematic practices of “not knowing”’ [2, 17]. An example of this was seen in other work by Webster when she took up how the evidence base for acute stroke treatment actually played out in the dayto- day work of the medical teams implementing what was considered best practice. She concluded that this EBM discourse, (amongst others), ‘designed to improve patient care come into view as managerial tools designed to control the delivery of care’ [32]. Of interest to all working in healthcare, she contrasted the text-based EBM discourse with the tensions experienced by physicians trying to implement ‘best practice’ on the ground, in their various work situations, organized through texts in the forms of guidelines. Similar tensions were experienced by paramedic, nurses, and dispatchers in Corman’s research on the social organization of emergency medical services [2]. There are multiple examples of disjuncture between policy and practice in both healthcare and HPE where institutional ethnography could be informative not only to explore, but also to inform meaningful change. One possible future exploration where institutional ethnography could be useful is the current vogue for the use of portfolios in the training of healthcare professionals, and the disjuncture between the differing reactions they bring about in faculty as they enforce them compared with the experience of students required to complete them to achieve a license to practice; not to mention the strict bureaucratic confines they all find themselves adhering to [33]. Inevitably connected to the aforementioned areas of potential study, we posit the need to critically examine the ‘student experience’ of becoming a health professional that deploys an institutional ethnographic lens. Studying the processes of becoming a healthcare professional such as a doctor, nurse, paramedic, or other allied health professional requires a nuanced approach that can move beyond the dominant focus on the lived experience of students to examine how such experiences are socially organized within a broader educational and healthcare system. However, research to date tends to focus on the perceptions or experiences of HPE students, or, at least in the context of learning to doctor, explores how medical ‘culture’ or ‘medical habitus’ is instilled during (and after) medical school. Examining the social organization of HPE from the standpoint of key players, such as students and their teachers, will greatly add to HPE as the social organization of HPE has yet to be explored. Finally, there is an increasing body of work in institutional ethnography specifically looking at the ‘health work’ that individuals do to navigate their health and healthcare (see earlier example) [6, 26, 28]. One of the authors of this article has a particular interest in the study of ‘high cost users’ and exploring how the health and social care system is organized to produce such trends. For instance, previous research in Canada has shown a consistent trend of a very small percentage of the population—often and problematically referred to as ‘frequent fliers’—accounting for the majority of public healthcare costs. Most studies to date on this topic have been driven by positivist epistemology, using primarily quantitative methodologies aimed at statistically understanding the demographics and drivers of this group. Little research to date has provided a complex understanding of this population and those who care for them. For example, Wise-Harris et al. [34] argue that, ‘the patient perspective’ of high emergency department users ‘is rarely represented’ in this domain of research. In addition, the perspective of those tasked with providing care to these supposed ‘frequent fliers’—their formal caregivers—are also left in abeyance in studies to date. There is a need to open up the unspoken realm of these individuals’ lived-experiences in accessing healthcare and social services, and begin to explicate how their experiences are organized to happen as they do. So perhaps adding to this literature on the ‘health work’ involved in being a patient can help give an authentic voice to patients and inform the training of those who will attend to patients in the future.

Conclusion

We wrote this article in light of our desire to elucidate institutional ethnography as a theory/methodology of discovery that has much to offer HPE and to encourage its use. We have argued ‘Why institutional ethnography? Why now?’ for HPE research. More specifically, we argued that institutional ethnography is well-suited to empirically examine a variety of topics of interest to HPE researchers, ranging from what HPE looks like in practice based on a diversity of standpoints (e. g. students, teachers, etc.) to the ever-changing institutional environment of HPE and how such changes are (re)organizing the work of those on the frontlines of HPE. We also suggested a strong empirical examination of how the ‘student experience’ is socially organized and the work (broadly conceived) involved in producing ‘competent’ students who will soon become practising healthcare professionals. In this light, we suggest institutional ethnography provides a critical and particularly informative approach to consider in moving HPE forward and encourage HPE researchers to make the ‘ontological shift’ we have described in Dorothy Smith’s innovative approach. We argue that institutional ethnography allows for a theoretically informed, albeit critical, lens to explore why there are large gaps between what is intended to happen (in policy) in HPE and what actually happens in practice. With that said, taking up institutional ethnography in HPE is likely to have its challenges as students and researchers conform to institutional requirements. Smith mentioned in conversation with us how she saw a strong potential for institutional ethnography in this domain but stated that ‘one of the things that I am concerned about is that I suspect that things like funding and the requirements of doing Theses, tends to detach the kind of connections with activism that were part of institutional ethnography’s origins, so it becomes something that is sort of ethnographic within a discipline but not related to the possibilities of making change’ [25]. Whilst we do not naively want to simplify the nuances of this very particular approach to inquiry, we will state that a less substantive background in the social sciences (all but one of this authorship are practising doctors) may mean that we are less burdened by the orthodoxy of those trained in more conventional sociological traditions. For those of us who have found our philosophical home in a less traditionally positivist paradigm, we are at ease with the ontology of institutional ethnography and reject its alternative sociology status. We muse that it actually seems quite intuitive to us as clinicians, reflecting our desire to maintain the subject as the subject and keep individual patients (and the students who will one day care for them) at the heart of our research, education and clinical practice. We suggest that many cultural shifts are beginning to align which has brought institutional ethnography to the fore as an informative and potentially transformative approach to inquiry with which to tackle many of the social issues in HPE as suggested above. In a time of increasing pressures on frontline health workers with emphasis on ‘accountability’ and ‘competence,’ institutional ethnography offers informed resistance to neoliberalism, which having already infiltrated healthcare practice and the institutional settings in which that practice occurs, threatens much of what we do as health professions educators. Anyone interested in the various tenets of how social justice transpires in HPE should give institutional ethnography some informed consideration. Funding The study fees and maintenance for this research comes from a Research Studentship from the Northern Ireland Department for the Economy (DfE). Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License

(http:// creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

The number of institutional repositories has been steadily building since the late 1980s and has rapidly expanded within the last decade. Research shows this rapid growth is dueto an increase in digital information, a growing awareness of open source publishing, and improvements to software (Asunka et al. 2011; Brown et al. 2004; Burchill et al. 2000; Chapman et al. 2009; Cox and Corrall 2013; Crow 2002; Graham 1995; Jean et al. 2011; John 2005; McGeachin 2010; Reilly 2013; Shearer 2002; Turner 2007; Westell 2006; Yiotis 2013). Collaborative programs and projects have been established nationwide and internationally to build, manage, and increase access to digital material in institutional repositories. Institutional repositories have had a very organic growth, meaning institutions can set prerequisites to suit their individual needs in lieu of conforming to set standards. Institutional repository programs have been developed by government bodies, academic institutions, and public entities. Different purposes of the institutional repository result in a variety of definitions, names, and plans. Despite the vast differences amongst institutional repositories, generalizations can be made. Crow (2002) defines institutional repository to be ‘‘stated broadly, digital institutional repository can be any collection of digital material hosted owned or controlled, or disseminated by a college or university, irrespective of purpose or provenance’’ (p. 3). In addition, Shearer (2002) noted that institutional repositories follow a set of characteristics, which are digital in nature, institutionally defined, scholarly, cumulative and perpetual, open access, and interoperable. Chapman et al. (2009) described, ‘‘an institutional repository as an entity that collects, manages, and disseminates materials produced at an institution’’ (p. 310). As with many developing areas, institutional repositories came in stages. The current state of institutional repositories focused on a macro approach to dealing with the number of issues surrounding the development and enhancement of the field. However, the history of institutional repositories is very much unknown in its entirety. A number of case studies have been conducted over the years, but in order to begin to move the field further, a temporal analysis of institutional repository research is needed. Fortunately, using data mining and information visualization techniques, a multitude of information on institutional repository studies can be gathered and analyzed to provide insight into the field. The study’s objectives are to understand the temporal changes of the field of institutional repositories, use information visualization techniques to discover subject themes on institutional repositories over four defined time periods, reveal relevant subject clusters, illustrate connections among related subject terms, and shed light on the history, development, and future studies of institutional repositories. The patterns revealed from the Chi Square analysis, multidimensional scaling (MDS) cluster analysis, and parallel coordinate analysis help to explain changing patterns of related topics to institutional repositories over time. In addition, the findings of the clustering analysis help researchers enrich the institutional repository related thesauri and develop subject directory systems of institutional repositories.

Literature review

Institutional repositories

Traditional repositories collect, catalog, preserve, and provide access to the collections. Institutional repositories accomplish the same tasks with the potential bonus of increased accessibility and collaborative efforts, which are created by building a sustainable program. One noticeable difference between traditional and institutional repositories is that collaborators are potentially the contributors. The big push for the development of institutional repositories began to take shape in the early 1990s when various academic disciplines began to contribute scholarly work through open-access publishing. Different disciplines and private industries began to experiment with collaborative projects in order to share information. In 1995, Graham used the term ‘‘digital research library’’ to provide possible answers to the growing volume of digital scholarly material. He called for collaboration between the university and library to contribute monetary means to maintain electronic resources while costs increased. The growing cost of published electronic material pushed different institutions and disciplines into discussions about the development of what is now typically described as an institutional repository. Graham (1995) stated the development of such a system is the mission of the library and should be viewed as an extension of the already established library system. Furthermore, Graham stated the characteristics of the digital research library should store, provide accessibility, and authenticate. For example, in 1996, the University of Illinois at Urbana-Champaign worked to develop the Digital Library Initiative, which provided full text retrieval to a number of digital repository collections (Schatz et al. 1996). During a time when digital material was becoming more apparent and difficult to efficiently deal with, more institutions began to experiment with similar digital projects, such as the University of Illinois at Urbana-Champaign. The early to mid-2000s brought a plethora of case studies from various institutions who had implemented institutional repositories. For example, Brown et al. (2004) noted that the automotive industry has seen an ‘‘increase in need for sharing and integrating information and knowledge between team members. An important aspect of this sharing and integration is the capture of key information and knowledge…’’ (p. 173). The development of the California Digital Library was founded to achieve, ‘‘permanent public access to local history object collections created by cultural institutions in California’’ (Turner 2007, p. 14). Each case study provides a window to many types of organization, software, and policy issues that were developed surrounding institutional repositories. After the mid-2000s, evaluation of institutional repositories began to occur more widely; this helped to propel the development of more institutional repositories and in part began to pave the way for case studies. Westell (2006) proposed a potential framework for evaluating institutional repositories and found that a better integration of academic researchers into the development of institutional repositories needed to occur. However, evaluating institutional repositories is a unique development, as they can each develop a program that works specifically for them. Institutional repositories provide institutions with the opportunity to share and preserve collections for a wide audience. The building blocks of development for an institutional repository are partnership, development of guidelines and principles, funding, software, collection of material, metadata, preservation, and access. These steps are not strictly sequential; more often, this is an iterative process in which each step is revisited many times throughout the development. A considerable amount of time, energy, and knowledge is required to establish an institutional repository through combining resources. Institutions began to realize that a collaborative effort to establish a digital repository could be achieved. John (2005) discussed a number of these issues, which included the field’s struggles and institutional repositories unsure future. However, the development of institutional repositories is such a unique experience for institutions that much of the knowledge has become a victim of the silo effect, meaning that information was circulated amongst researchers and academics most closely associated with a single institutional repository.

Temporal data mining

Data mining techniques can be used to retrieve information created about institutional repositories from the inception. Temporal data mining provides a means to observe patterns that have occurred throughout the years and gather large amounts of data about the subject terms that have been used to describe institutional repository research. Data mining is viewed as a highly robust method of research. More recently, temporal data mining has become an important method to identify large pattern changes over a set course of time. Fernandes (2014) used data mining as a way to identify trends within software engineering; a major finding included how software engineers have been collaborating in more recent years. Major factors of temporal data mining are the value of time and the ability to identify patterns and categories within the data gathered. Esling and Agon (2012) found the involvement of time series data mining enabled representation, distance, and indexing methods of a particular research area. Atunes and Oliveira (2001) described a number of essential techniques and methods that can be applied during temporal data mining. For example, essential techniques include the discovery of relations to occur during data modeling, definition of similarity measures, and application of models and representations. Keyword data mining is another avenue for researchers, especially when in conjunction of citation analysis (Wang et al. 2013). A variety of research can be conducted through the incorporation of data mining techniques. Au Yeung and Jatowt (2011) experimented with large-scale text mining and time compressions to research large datasets of news articles to enhance cultural history. Alonso et al. (2009) found that time is particularly important when seeking the measurement of a particular period in order to provide the proper context for retrieved documents. Recent studies involving eye-tracking systems had also incorporated temporal data mining techniques. Yu et al. (2012) used visual data mining to examine patterns of knowledge discovery in early human development. Institutional repositories had an organic development, meaning that while collaborations between institutions were being established, there was no substantial tracking of overall development. Since the initial development of institutional repositories in the 1990s and 2000s, there has been a significant amount of technological advancement, which has changed how people use, manage, and find information. As a result, an analysis of information would be useful in determining patterns and differences. Mehta and Dang (2011) noted that the field of temporal data mining is expanding quickly due to the amount of temporal data stored and a recognition regarding the importance of temporal data. Temporal data mining should then be applied to institutional repositories. However, another important aspect of data mining is information visualization. It is with visualization method that the patterns discovered in data mining can be displayed.

Information visualization

Information visualization and clustering analysis demonstrates relationships among data. These relationships can be particularly useful when combined with longitudinal or temporal data. There have been a number of bibliometric studies that focused on keyword conetworks in order to ascertain the trajectory of different fields (Liu et al. (2015). Ke (2013) used information visualization to observe citation modeling from InfoVis 2004 data, and was able to demonstrate through strong visualizations, the changes of the field. In a similar study, Chen (1999) was able to demonstrate semantic structures and co-citation networks of digital libraries using visualization techniques, particularly the grouping of similar data. Through this method, visual perception is an aspect that aids in the process of discovery and exploration of data sets. Several studies have used visualization techniques to gather information about subject term analysis. For example, visualization methods have been used in information retrieval studies to discover more about user behavior and associations between subject terms in a given research area (Zhang et al. 2008). Wang et al. (2003) noted that a query clustering at the conceptual level could provide a substantial amount of data for system development. Lee et al. (2012) used visualization and clustering methods to demonstrate clustering of large data sets. In relation to this study, connections were instead made by an analysis of the subject terms lending to information regarding the development of institutional repositories. Robertson et al. (1989) developed a framework describing a visual application that would potentially be used to represent information and its structural relationships in a 3D environment. This visual application can be used to draw out relationships that are not inherently evident. Chen et al. (2004) proposed a visualization tool that provides insight into the complexities of analysis. The framework provided by Chen et al. (2004) demonstrated that by using standard algorithms for visualization and discovery, visualization objects, and system architecture, a powerful starting point for research could be envisioned. Thus concluding, there is a great potential for institutional repositories. Large sets of data gathered through data mining techniques can be more easily conceptualized when combined with information visualization techniques and procedures.

Parallel coordinate analysis

Parallel coordinate analysis is a method that can be applied to discover data patterns. Parallel coordinates have been proven to be an effective tool when applied to data sets. Multiple theme threads identified from data sets can be revealed through coloration tools within statistical programming software. Parallel coordinate analysis offers the ability to visually display information from multiple viewpoints. For example, information can be categorized and then plotted to facilitate comparison and highlight items of interest (Heer and Shneiderman 2012). There is a variety of ways that parallel coordinate analysis is currently being utilized in research. In an effort to reduce clutter within the patterns found in parallel coordinate analysis, Zhou et al. (2008) optimized curves through color enhancement to improve visual effectiveness. Dasgupta et al. (2012) studied visual representations of parallel coordinate analysis and proposed the inclusion of a variety of taxonomy measures decrease the level of visual uncertainty, such as the use of clustering or MDS. Lee et al. (2012) used parallel coordinates view to illustrate the interaction and overview of document-topic distribution, which allowed users to have a more thorough understanding of the analysis. When applied to institutional repositories, these techniques enable the visualization of term increases. Furthermore, unchanged trends change to provide an overview of the field.

Multidimensional scaling

MDS is another method of information visualization that demonstrates similar patterns across data sets. MDS data demonstrates to a large degree the different levels and patterns that have changed over the years as institutional repositories have matured. Visualization large-scale MDS data can display new connections and patterns that have evolved over the years (Dzemyda et al. 2011). Zhang et al. (2008) used MDS to analyze medical topic queries to demonstrate relationships among query terms. The MDS information visualization method provides an important avenue into the interaction that has occurred within certain data sets. Zhang and Meng (2013) used MDS to reveal the semantic relationships amongst subject terms related to the metadata standards of Dublin Core research. It is important to use MDS for researching institutional repositories because a large number of subject terms are then clustered to provide an overview of the evolution of the field.

Methods

A pilot study conducted prior to the study fulfilled three distinct purposes. The first finding of the pilot study established a complete list of search terms related to the topic to ensure the completeness of the collected data. 328 unique terms were identified on the topic of IR. The second purpose of the pilot study was to determine the year of the earliest publications related to IR and to define the periods for data analysis. Ideally and in principal, the defined periods should be equal in length, and each period should have sufficient data for data analysis. However, the pilot study determined that earliest year related to IR was 1992. From 1992 to 2001, there were few studies published during the time of 1992–2001. However, following 2001, many more studies were published on IR. Each period should be efficient enough to include enough data; the early periods were very limited with publications. Consequently, the first period covers a longer time span than subsequent periods. The average is of the later periods is 3 years. The third and final purpose of the pilot study was to generate an initial schema, which laid a foundation for the later content and data analysis. The schemas included patterns revealed from the collected keywords, for instance, theoretical frameworks, software development, and management. Toward the aims of this study, the following hypotheses were introduced.

H10 There are no significant differences of the number of subject terms within the category Theory and Methods across the four time periods.

H20 There are no significant differences of the number of subject terms within the category Application and Management across the four time periods.

H30 There are no significant differences of the number of subject terms within the category Implementation and Technique across the four time periods.

The presented hypotheses attempted to ascertain the term changes within each of the three mega-categories over the four defined periods. The definitions and justifications of the mega-categories and time periods are addressed later. Since category was defined as the dependent variable in each of the presented hypotheses, it is measured by its frequency. A Chi Square for goodness of fit was used in the presented hypotheses because each of the hypotheses involves one dependent variable and the primary purpose of the hypotheses is to examine whether the number of terms that falls in each of the categories differ significantly from the expected number. In each of these tests, the expected frequency percentage for each of the four periods is equal to 25 %. The significance level for the tests is set to 0.05. If a presented hypothesis is rejected, it indicates that subject terms in that specific category changed significantly over the four periods.

The aim of the hypothesis is to uncover how the terms changed over time within each mega-category. The focus of the paper is to derive how institutional repositories have developed and evolved as studied in published academic works. In conjunction, with MDS and parallel coordinate analysis, Chi Square was used in order to thoroughly examine the term changes across the four time periods within each category. The hypothesis focus on the particular terms in each mega-category. MDS and parallel coordinate analysis methods were used to analyze subject terms of published research about institutional repositories in order to demonstrate how the field has developed. The MDS method focused on the subject term relationships within each of the four time periods, while the parallel coordinate analysis method concentrated on the subject terms’ connections across the four periods of time. However, the analysis of the subject term changes, the study was able to identify term decreases and increases over time. The identification of term change trends on institutional repositories gives particular insight into the evolution of institutional repositories and sheds light on where the field is going.

Selection of databases and data collection

Data was collected from the Web of Science (WoS) citation databases to conduct data mining analysis. The WoS databases provide access to numerous citations in a number of different disciplines. Several different databases are accessible through WoS and have over 100 years of abstracts, 12,000 journals, and 148,000 conference proceedings in 250 disciplines. As a result, WoS provided an ideal investigation environment to find the highest level of research conducted on institutional repositories (‘‘Web of Science,’’ 2015). Research about institutional repositories was extracted from WoS. The data collection procedure included a number of chronological steps to ensure all information regarding institutional repositories would be gathered. First, a series of subject terms were identified to perform exhaustive searches within the WoS databases. The queries included a number of different terms to extract information from the databases. The lack of a unanimous definition of institutional repositories, in addition to its interdisciplinary nature, required that a number of different queries be conducted to perform successful data collection. For example:

• Institutional repository

• Digital repository

• Information repository

• Open repository

• Digital institutional repository

• Data repository

• Digital research library

• Open archive

• Open access initiative

Queries were entered in the ‘topic’ field in the WoS databases. The topic field provided a meaningful selection of subject terms and then allowed the terms to be entered into the WoS databases in a way that examined the root development of institutional repositories. Together, these terms resulted in the successful extraction of institutional repository data. Since terms from the title, abstract, and subject fields of a study characterize its subject topics, they were used to analyze topic changes. The mega categories resulted from a coding analysis based on keywords extracted from the investigated documents. In other words, these categories are emerging themes from the documents on institutional repositories found in the database. The documents were collected based on keyword extraction from the document titles, abstracts, and through searching the full text of the document. From these collected keywords, the authors were able to determine emerging themes from the collected document literature. The extraction of subject terms from the titles, subjects, and abstracts of research publications, provided a richer array of data for the study. The inclusion of subject terms in the abstracts and titles also ensured the publications were, in fact, about institutional repositories. Examining the publication abstracts allowed for a high level of continuity and consistency throughout the study. For example, the meaning of certain subject terms like data repository may appear as though it would fit within the directory of institutional repositories; however, this is not the case. In order to ensure that the subject terms extracted were associated with institutional repositories, even full-texts had to be reviewed. Once the subject terms were extracted from each study, ‘like’ subject terms were combined in order to decrease accuracy. For example, terms like institutional repository and institutional repositories were combined. Stop words, such as, how, what, and at, were not included in the collected data. Subject terms used once were removed from the data set because they had little contribution to the later term cluster analysis. The terms came from the title, keyword, and abstract fields of the bibliographic records in the database. The majority of abstracted keywords are nouns. The extracted keywords were validated. For instance, the plural forms were changed to their singular forms. Meaningful adjectives were converted to their nouns. Meaningful verbs were also converted to their nouns. In addition, synonyms were merged to represent one like noun. A Kappa agreement test was conducted to evaluate the consistency of the two coders. The resulting Kappa value is .759. It implies that there was a substantial agreement between the two coders according to the criteria (Viera and Garrett 2005.)

Determination of time periods

The interdisciplinary and dynamic nature of the institutional repository study, as it involves many subjects, shows that some phase out, and some appear; it is necessary to investigate the subject change in the dynamic period through the lens of the time. A pilot study revealed there were four distinct periods of developmental changes with institutional repositories. Four time periods were identified using the temporal analysis method during the course of this study. Consequently, the study was divided into four distinct time periods.

• Period I: 1992–2001

• Period II: 2002–2005

• Period III: 2006–2009

• Period IV: 2010–2013

Temporal analysis is suitable for a topic that is a relatively new, emerging research area, where research on the subject is not fully mature. This characteristic fits the topic of institutional repositories. The purpose of the temporal analysis was to gather institutional repository data to provide a rich array of material to discover developments and changes over time in the field. Toward this aim, this study made a clear division of the four time periods. Each period, other than the first, had an equal division of years. However, Period I, 1992–2001, gathered the least amount of research as compared to the other periods. Period I also has the largest number of years associated with it, which was due to the studies being produced at the time. A greater number of years in Period I provided a larger number of documents to be gathered, which were of closer relation in regards to later periods. The last three periods in this study were divided by 3 years. Upon evaluation of the data collected, it was clear that a 3-year division between Periods II, III, and IV provided the most stratified sets of documents. In doing so, the period divisions provided a rich source dataset to be analyzed for the study. Another pilot study found that research concerning institutional repositories had been gravitated toward three distinctive themes. These themes were defined as three megacategories. The three mega-categories allowed the researchers to focus on term changes over the four periods of time.

Category 1 Theory and Methods

Category 2 Application and Management

Category 3 Implementation and Technique

Table 1 illustrates the subject terms that were assigned to each mega category as well as the corresponding documents that are associated with them. In the Table, PI–PIV stand for ‘Period I, Period II etc. In Table 1 the unique terms increased dramatically for Implication and Technique from periods I to period IV. This is an indication of how the subject area that new concepts have been developed like new technology and methods have been added during this period of time. The number of documents have also increased, this suggests that more studies have focused on the topic. In contrast, Theory and Methods has had an increase in the number of documents, but has had a stable number of unique terms over time. This is an indication that more research is taking place within the field of institutional repositories about Theory and Methods, but the stable number of unique terms indicates that few new theories or methods are being developed.

Multidimensional scaling method

With the MDS analysis, the study was able to identify terminology changes within a particular period for institutional repositories. Each period indicated changes as the field emerged and developed. There is a large variety of mature information visualization techniques, such as Pathfinder associative networks, self-organizing maps, MDS analysis, and so on. MDS was selected because the requirement for the processed data is free of any distributional assumption. In regards to the involved data, there is no restriction on data scale, and it fits a small- and middle-size data set (Zhang et al. 2008). Information visualization techniques such as MDS can efficiently convert sophisticated and abstract object relationships in highdimensional space into a low-dimensional space, where objects’ semantic relationships can be visually analyzed and examined. As a result, the subject terms from each of the four periods can be projected onto the low-dimensional space where relevant terms are grouped and clustered, and their relationships can be revealed and analyzed. The stress value is an indicator of MDS to ensure the quality of the visualization result is properly clustered. The stress value shows that the smaller the stress value, the more accurate the result (Zhang et al. 2008). If a resultant stress value is below 0.15 or r2 (RSQ) value is over .90, the corresponding MDS result is considered acceptable. SPSS was chosen to create the MDS results, and the stress value was measured using the Euclidean distance model. Terms obtained from the databases were extracted from fields of records and validated to generate a term master file. The record-term matrix within each of the four periods was created based on the term master file. Then, the record-term matrix was converted to a term–term matrix based on the Cosine similarity method. Finally, the term–term matrix served as input raw data for later MDS’ analyses. It is recognized that a document of interest may be multi-faceted in terms of its content, meaning that it is normal for a document to cover multiple subjects due to interdisciplinary nature or the complexity of a research topic like institutional repositories. These subjects are represented and indexed by individual keywords in the database after subject content analysis, and each keyword from the document that represents one of the multifaceted meanings of the document. In the coding analysis process, the deconstructed keywords rather than the documents are classified into one of the mega-categories. Following the definition of the four periods, the investigated documents were divided into four periods. The documents were then processed separately, and keywords extracted from the documents in the different periods were tallied separately. Consequently, the intensity of the same keyword varies in different periods and in different mega-categories as well.

Temporal analysis

In order to gauge the development of institutional repositories, this study applied the parallel coordinate analysis method to provide a richer description of the subject term trends over time. The purpose of applying the parallel coordinate analysis was to discover term increases, term decreases, and unchanged term trends over the defined four periods. Through this kind of term analysis, a deeper understanding of institutional repositories was able to be conceptualized. XLSTAT was used to visualize the data (‘‘XLSTAT,’’ 2014). The subject terms gathered in this study needed to be applied to a larger categorical scheme in order to make sense out of the data. Consequently, the data was organized into three mega-categories: Category I (Theory and Methods), Category II (Application and Management), and Category III (Implementation and Technique). From this point, the subject terms from each period were separated into the three megacategories. The terms were placed into the category that was the best fit for the subject matter. For example, subject terms like discipline and information science and library science were placed in Category I (Theory and Methods). Subject terms placed in Category II (Application and Management) were knowledge management and Dspace. Subject terms found in Category III (Implementation and Technique) had subject terms like experimentation and dissertation. The goal of this analysis was to analyze the term changes used in institutional repository research to discover the developments of the area. The parallel coordinate analysis was then organized into four figures. The first figure visualized the overall term changes across the defined four periods. The second figure displayed term change in Category I: Theory and Methods across the four periods; the third figure demonstrated term change in Category II: Application and Management across the four periods; the fourth and final figure illustrated term change in Category III: Implementation and Technique across the four periods. With parallel coordinate analysis, the study was able to identify terminology changes over time for institutional repositories. The parallel coordinate analysis also showed the relationships about how different disciplines began to work with institutional repositories.

Results

The results yielded 316 records that were collected from 1992 to 2013. The initial total number of keywords extracted was 1065; these keywords were the raw terms. Many of the original keyword terms were combined to form subject terms. In particular, the original keywords that were duplicates or were similar (institutional repository and institutional repositories) were combined, which left the total number of the validated subject terms at 230. The distribution across the four periods for all the subject terms is located in Table 2. The minimum number of keywords was found in Period I with 39; Period II had 52, Period III had 68, and Period IV consisted of 77 terms. The maximum number of terms was 77 found in Period IV. Overall, the average number of subject terms was 57.5. The most frequently used subject terms per period can be found in Table 3. General statistical information regarding the subject terms that was gathered during the study can be found in Table 4, which shows the mean, standard deviation, and number of records that were gathered from WoS databases. For each period, subject term clusters are visually displayed in the following figures that follow each narrative description. Resultant p values for the hypotheses H10, H20, and H30 are 0.852, 0.142, and 0.000 respectively. The test results suggest that H10 and H20 were accepted because their corresponding p values are larger than 0.05. Moreover, H30 was rejected because its p value is smaller than 0.05. In other words, there are no significant differences of the number of subject terms within the category Theory and Methods across the four time periods at the significance level 0.05; there are no significant differences of subject terms within the category Application and Management across the four time periods at the significance level 0.05; and there are significant differences of subject terms within the category Implementation and Technique across the four time periods at the significance level 0.05. The detailed test results are shown in Table 5. It is clear that the test results of the presented hypotheses can only demonstrate whether there are significant differences of the subject terms in the category across the four periods. They cannot tell how the subject terms changed within a category, what subject terms caused the changes, or what the relationships of the term changes across the four periods are. The following MDS analysis method and parallel coordinate analysis method can effectively answer these important questions.

Part I multidimensional scaling results

The following are the results of the MDS analysis. Outcomes are summarized by a narrative analysis and followed by a visual display of the terms clustered, which are illustrated in Figs. 1, 2, 3 and 4. In addition, the term clustering results are demonstrated in Tables 5, 6, 7 and 8. In each table, the terms were labeled. Labels were used to maintain the clear visual structure without the visualization becoming too cluttered from using full subject terms. For example, in Period I, {T13; T16; T30; T12} are all found in Cluster 1. The tables illustrate the full narrative for each T number: T13 (Hardware and architecture); T16 (Information science); T30 (Multidisciplinary sciences; and T12 (Engineering).

Cluster analysis from 1992 to 2001

Period I (1992–2001) had five identifiable subject clusters. The stress value for Period I was .10694 and r2 was .95621. The subject terms associated with each cluster can be categorized into the following areas: software and technique, theory and methods, discipline, institutional repository, and development. The visual display of the terms in this period is shown in Fig. 1 and term clustering results are demonstrated in Table 6. Period I yielded the fewest results of all periods researched (39). The low number of subject terms from this period is due to the low number of terms that were primarily associated to the studies from the computer science discipline. A possible explanation for the low number of terms associated with the research is that during this first period, institutional repositories were at the beginning of development. Many of the studies produced during this period were technical in nature. Cluster 1 connected five different subject terms. These terms were associated with research that discussed the possibility of different applications and software platforms that could be employed by institutional repositories. The association of the term agriculture was associated with software development research. This was due in part to the large number of studies already available from the discipline of agriculture that had begun to develop institutional repositories in the 1990s. For example, Common Object Request Broker Architecture (CORBA) was the first term associated with a sharing data structure for research. CORBA was an early example of what would become known as institutional repositories. Interestingly enough, research discussing CORBA noted that due to the complexities associated with databases and sharing infrastructure, more work would be necessary to establish a strong application to be utilized in a sharing environment (Beck 2001). Cluster 2 joined terms that dealt with theory and methods within institutional repositories. The importance of theory and methods within Period I was evident from the research that was being conducted. Much of the research contained a great deal of discussion amongst those in different fields, though the largest developments were from those with computer science backgrounds. Cluster 2 was also the largest clusters of terms from Period I. This also provided evidence to the amount of discussion that was being conducted in the 1990s on how institutional repositories should be developed and managed. Studies from the information science and library science field produced research with suggestions about where and how institutional repositories might be useful in various environments. For example, a noticeable amount of research was concerned with the organization, metadata development, and the role of libraries within the development of institutional repositories (Bertino et al. 2001; Burchill et al. 2000; Casas et al. 2000; Dempsey 2000; Graham 1995; Hedstrom 1998; Helly et al. 1999; Holmes and Mathieson 1999; Karvounarakis and Kapidakis 2000; Smith et al. 2001; Vetterli et al. 2000; Wilensky 1996; Wu and Liu 2001). Cluster 3 focused primarily on the different disciplines that could be affected, as well as who should be developing institutional repositories. Cluster 3 was one of the smaller clusters in Period I, which resulted due to a lack of research produced during 1992–2001 that focused specifically on technological aspects within institutional repositories. The field was just beginning to emerge in the early 1990s, and more of the focus was on general development, rather than which field or fields should be harnessing institutional repositories. Research was more focused on theory and development instead of the narrowing of one specific software application or platform. Cluster 4 focused on research development in the very early stages. The key make-up of institutional repositories is that research is accumulated into one space and made available online for researchers of various disciplines. Cluster 4 was an indicator of this exact characteristic. The key terms from Cluster 4 demonstrated a strong focus to develop definitions and characteristics on institutional repositories. In addition, a noticeable portion of the research remained theory-based due to the emerging nature of the field at the time. Interestingly, Cluster 4 was closely related to Cluster 2 (technology); this factor demonstrated that the emphasis during Period I was on what type of technology platforms would be best utilized for institutional repositories as the field moved forward. This cluster finding provided more evidence on the developing nature of institutional repositories in Period I. For example, a few terms that appeared in Cluster 2 were computer science, software engineering, repository, and theory and methods. In contrast, information science and library science was a term that was associated with Cluster 4; this was also an early indicator of where institutional repository research was moving. Cluster 5 focused on development and was very closely aligned with Cluster 4 (research and development). The strongest focus of the terms associated together in Cluster 5 had a general indication of development in different respects. Some of the research associated with these terms varied from developmental standpoints, but had a theoretical basis, while others varied from technology to development, and others from general management to development. For example, Plante et al. (1999), Purvis et al. (2001), and Li et al. (1999), all conducted research that explored many different levels of innovative design and/or innovative ways of using at the base level, which is referred to in this study as an ‘institutional repository.’

Cluster analysis from 2002 to 2005

The results for entering the search term ‘institutional repository’ during Period II (2002–2005) indicated a number of characteristics. The stress value for Period II was .13375 and r2 was .90538.

Period II (2002–2005) had three identifiable subject clusters: development, technique, and policy. The visual display of the terms in this period is shown in Fig. 2, and term clustering results are found in Table 7. Period II grew in the number of terms by 63 % as compared to Period I. The subject term data repository was common in studies relating to health care and information retrieval. While Period I discussed a wide array of possibilities and potential areas of development for institutional repositories, Period II focused on how they were a good idea, yet lacked technology and support.

Cluster 1 connected ten different terms that all dealt with development issues associated with institutional repositories. As opposed to the first period, Period II’s research was becoming more focused on specifics of institutional repositories but was still very much in the experimental phase of development with a focus on future planning. The terms conceptual knowledge and semantic web were evidence of the future research focus. Other broad terms such as consortium, communication, and community also provide insight into the types of professional and research groups that would use institutional repositories.

Therefore as previously stated, Period II represented development of institutional repositories as a whole with a focus on future growth. Cluster 1 was also closely related to Cluster 2. Cluster 1 also covered the widest breadth of study on Period II. Within the time period of 2002–2005, the term development was being explored in a number of different ways and perspectives. For example, the terms digital repository and national were both associated with research that tried to address issues dealing with what types of communities and digital collections would be best maintained in institutional repositories.

Cluster 2 focused primarily on techniques within the scheme of institutional repositories. The association of technique within this cluster was drawn from research that discussed different possibilities about how the institutional repository could be structured and who would be the users of the repository. As institutional repositories were still in development at this time, the research was very open with discussing what techniques might be best utilized for future planning and structuring of systems. The association among subject terms such as open source, metadata, and open content license were evidence of the technique trend.

Cluster 3 represented related terms that dealt with policy. In Period I, Cluster 3 was associated with disciplines and how each discussed institutional repositories that were drawn from research, as well as who should be developing institutional repositories. Period II, Cluster 3 was the next stage of that discussion. In this period, Cluster 3 represented the extension of development that was able to be achieved from Period I to Period II. For example, Period I, Cluster 3 contained research that discussed possible services for digital research libraries (Crespo and Garcia-Molina 1997). In Period I, Cluster 3 also discussed management through harnessing web tools and sources (Karvounarakis and Kapidakis 2000). Moving forward, Period II, Cluster 3 represented subject terms that were associated with research that reviewed metadata within the context of institutional repositories, collaboration research, and potential design structures and frameworks of institutional repositories (Schweik et al. 2005; El Demerdash et al. 2005; Lee and Yoo 2005; Cui et al. 2005). The research from Cluster 3 emphasized research studies conducted on institutional repositories and the field’s future development.

Other subject terms in Cluster 3 specifically focused on theory and the practicalities of institutional repositories. For example, of the 35 studies in 2005, nearly all of them focused on specific repositories and systems. Terms such as publication, international, access, and open access were intertwined with studies that addressed the multitude of issues associated with institutional repositories and scholarly publishing. As depicted of research conducted during Period II, no definitive answers were provided to many of the policy questions raised. Instead, suggestions for future research in the area were proposed. For example, John (2005) and Yiotis (2013) both raised questions about open-access issues and were concerned with the future growth of institutional repositories. The suggestions for future potential guidelines or standards were a noticeable element of Cluster 3, especially with the terms that were associated with research in 2005, as much of the research was looking ahead toward future development.

Cluster analysis from 2006 to 2009

The results for Period III (2006–2009) indicated a number of characteristics. The stress value for Period III was .13682 and r2 was .92825. Period III had 68 subject terms associated with the time span. Period III (2006–2009) had three identifiable subject clusters: design, collaboration and policy, and institutional repository. The visual display of the terms in this period is shown in Fig. 3, and term clustering results are demonstrated in Table 8.

Unlike Period II, Period III focused more specifically on institutional repositories and began the process of developing ideas and discussion around the expansion of institutional repositories. Consequently, a significant amount of research was conducted on case studies in this period.

Cluster 1 joined ten different terms together, which focused on design elements of institutional repositories. Different design perspectives were provided in research from different disciplines. There was also a strong link between the terms collection development, case study, and communication. Amongst the case studies, there was a strong connection between the specific design and the specific case. For example, within the element of design, the case studies discussed the different information systems and management techniques that were established. A few case studies specifically involved the user side of institutional repositories. For instance, Devakos (2006) explored user responses to institutional repositories at a case-study level; Munkvold et al. (2006) researched the decisionmaking process and the collaboration side of institutional repositories. Other case studies were more traditional. For instance, Palmer et al. (2008) reviewed three different institutional repository initiatives. Choudhury (2008) conducted a case study at John Hopkins on data curation at a university level. Cluster 1 demonstrated the growth in studies from Period II by discussing the actual development and uses of various institutional and digital repositories.

Cluster 2 coupled terms that dealt with collaboration and policy elements of institutional repositories. Cluster 2 was also the largest cluster in Period III, and it covered the most breadth of material. In Period II, the topic area of collaboration and policy was just beginning to become focused enough to provide future ideas and possibilities. The additional case studies that were researched and reported on in 2006–2009 provided a number of examples from which to draw new ideas. The fourteen terms in Cluster 2 articulated ideas discussed in the research that raised issues of intellectual capital, resource sharing, and international colleagues. The only discipline that was focused within this cluster was information science and library science. The majority of the ideas raised in information science and library science research were concerned with the long-term impacts of resource sharing and information management. Overall, the terms in Cluster 3 were closely associated with one another. It was not one cluster or relationship individually, but as a whole the cluster was able to include a plethora of information. Consequently, the terms within Cluster 3 were best represented by areas within research development. A number of disciplines were evident, and different software systems were discussed in association with institutional repositories. Some of the terms within certain areas focused more narrowly on the idea, like user interface, DSpace, and open source software. Cluster 3 was the best representation of what later emerged in Period IV as research development. In addition, research related to open source software and technological advancement for institutional repositories was a common thread in Cluster 3. For example, potential open source software models discussed possible ways to integrate new software systems (Afshari and Jones 2007; Tramullas and Picazo 2006). The subject term, DSpace, was of particular interest because of its role within the context of the implementation environment. Embedded into other case studies were also found in Cluster 3, and case studies within this cluster had a focus on technology systems and interface design (Bramscher and Butler 2006; Chen and Hsiang 2009; de-la Vega-Sivera 2010). Cluster 3 also associated subject terms that dealt with academic relationships. For example, publisher, academic author, and academic publishing were closely connected. The discussion among the research provided background to the possible impact of publishing and the relationship to institutional repositories. The discussions provided information on potential long-term contributors, who would be accessing the information, and where the repository should be located. More specifically, the location refers to the institutional repository home page site. The research provided an opening for technicalities to be discussed continuously as policies changed. The results found in Period IV indicated a number of characteristics for institutional repositories. The stress value for Period IV was .15266 and r2 was .91639. Period IV (2010–2013) had three identifiable clusters: publication, management, and institutional repository. The visual display of the terms in this period is shown in Fig. 4, and term clustering results are demonstrated in Table 9. Period IV culminated the expansion and development of institutional repositories beginning in 1992 through 2013. Out of all four periods, Period IV had the largest number of subject terms at 77. In Period IV, many of the trends that began in earlier periods were shown to be more mature. In addition, starting in 2010, research began to discuss how institutional repositories were coming into their own and what steps should be taken to advance them. Cluster 1 joined terms associated with areas such as publication and academic use. Within the context of institutional repositories, ‘collection’ refers to the items held in the repository. In addition, academic use refers to the shift in the use of terms like ‘publication’ within academia. The varieties of collections differ depending on the premise or mission of the institution. The terms dissertation, electronic thesis, and professional were variances between individual, institutional repositories, while some repositories could obtain all of the listed items and more. Evaluation was another term in Cluster 1 that aided in the establishment of institutional repositories, and the next step of evaluation has begun for several cases. Cluster 2 linked together terms that were associated with management, more specifically subject areas related to contribution and contributors to institutional repositories. Period IV also brought to light a number of well-developed ideas. In 2010–2013, many ideas related to institutional repository management were articulated by not only the users of the institutional repositories, but also the creators of the material. As a result, the cyclical process that occurs between contributors and users has the ability to change significantly due to the nature of institutional repositories. For example, faculty and communication were closely associated terms. Research associated with these terms analyzed open-access possibilities. Abrizah et al. (2010) researched the characteristics among Asian university institutional repositories. Asunka et al. (2011) focused on the possible integration of social networking tools to aid in communication among faculty involved with institutional repositories. The relationship in the research depicted how it was necessary within academic settings for the contributor (often faculty members) to communicate their needs as researchers and as users in order to assure the structure of the institutional repository was efficient. Cluster 3 associated subject terms that referred to variances of development and growth. Cluster 3 was also the smallest cluster in Period IV. Cluster 3 consisted of terms that were associated together by management. Management has evolved throughout the three periods of institutional repositories. In addition, research from Period IV came from the case studies that had been conducted in 2001. Consequently, research concerning management in Period IV consisted of a wide array of information from recommendations, case studies, and practices that had failed to succeed. A new term, business model, was an addition that demonstrated evaluation research was at the cusp of being conducted. In addition, user studies concerning institutional repositories were being treated much in the same fashion as academic libraries within higher education. Terms like continuum, open system, and users of IR (information retrieval) all alluded to the next phase of development in research and evaluation. Period IV was the beginning of the next phase of institutional repository research. Several studies associated in Cluster 3 demonstrated a variety of evaluation tools for institutional repositories. For example, Galina Russell (2011) focused on the critical view of institutional repositories and the association with open access. Other research dove deeper into areas of user perspectives and the metadata standards within the context of institutional repositories as a long term information source (Ivanovic et al. 2012; Kumar 2012; Sawant 2012). Shown in Table 10 is the summary of the stress values, r2 values, and the number of terms for the MDS results. The resultant stress values and r2 values indicate that all the resultant clusters from the MDS are sound. Each of the clusters in each period had subject terms all related to a specific subject. In Period I, Cluster 1 had subject terms most closely associated with areas that included software and technique. Period I, Cluster 2 had subject terms related to different theories and methods associated with institutional repositories. Period I, Cluster 3 had subject terms linked with issues that dealt with disciplines within academia. Period I, Cluster 4 had subject terms closely associated with issues of management. Period II, Cluster 1 had subject terms most closely related with the subject term, development. Period II, Cluster 2 had subject terms connected with the subject term, technique. Period II, Cluster 3 had subject terms linked with the subject term, policy. Period III, Cluster 1 had subject terms correlated with design. Period III, Cluster 2 had subjects coupled with issues that dealt with collaboration and policy. Period III, Cluster 3 had subject terms most closely associated with research and development. Period IV, Cluster 1 had subject terms that connected with publication. Period IV, Cluster 2 had subject terms related with management. Period IV, Cluster 3 had subject terms most closely associated with areas of development and growth. Table 11 shows the clusters with their associated subject areas.

Part II parallel coordinate analysis results

The parallel coordinate analysis expanded upon the MDS findings. If the MDS analyses focused on each of the four periods, then the parallel coordinate analyses concentrated on the term connection across the four periods. Through the categorization of the subject terms across the four time periods, a deeper understanding of the developments of institutional repositories was illustrated. A number of characteristics were drawn from the results of the parallel coordinate analysis, particularly the changing terms. The term changes through the different mega-categories demonstrated, as a whole, how the field of institutional repositories developed foundationally and how the research area matured. The disappearance, reoccurrence, consistency, and newly emerged terms all told stories about the field of institutional repositories. The parallel coordinate analysis results are visually displayed in the following four figures. Figure 5 illustrates the overall term changes across the four periods. Figures 6, 7 and 8 illustrates each mega-category change pattern using parallel coordinate analysis. Figures 6, 7 and 8 also indicate the terms in each mega-category throughout the four periods. Table 12 illustrates the most frequently used subject terms for each mega-category. The top five subject terms are listed with their frequencies in each mega-category. Table 13 displays the distribution of the unique terms and the number of records as they were divided amongst the three mega-categories.

Overall term changes

Figure 5 demonstrates the dramatic topic term changes that occurred from 1992 to 2013 for institutional repository research. The X-axis stands for the mega-categories, and the Y-axis stands for the time periods. Each period is illustrated with a different color. Figure 5 illustrates that Implementation and Technique had to highest number of terms associated with the mega-category in Period IV, whereas, Periods I and II had the lowest number of terms for Period IV. This finding demonstrates that institutional repositories had not yet refined implementation strategies by 2010, therefore, publications on the subject rose dramatically.

In addition, Fig. 5 Applications and Management had the largest spike in Period III. This finding indicates that it took institutional repositories a decade of research and development to find software and strategies that worked. This finding is also reflective of the technology boom as software and equipment became more affordable. Theory and Methods is also reflective of its time periods of importance. Period I has a strong association of research regarding theory and methods of institutional repositories, but as the foundational points had been laid, more work began in other areas like Applications and Management.

Theory and methods

Overall, the category Theory and Methods (found in Fig. 6) had subject terms in all four periods that rapidly appeared and then disappeared. Subject terms like conceptual knowledge, design education, and intellectual capital, all provide insight into the research that was being conducted. Figure 6 captures the rapid emergence of the terms that readily disappeared and then reappeared once more research had been conducted within institutional repositories, which foundationally structured the field to allow the category Theory and Methods to emerge. The average of term appearances in each period increased incrementally in each period, 2.42, 3.35, 4.73, and 4.54, respectively. This finding further strengthens the findings found using MDS. For example, the MDS results indicated that a large amount of research being conducted during the inception of institutional repositories dealt with computer science and related disciplines (i.e. software engineering). The field then later changed to be more influenced by the most-used terms in the mega-category, which was information science and library science, with the term appearing in Period I– Period IV as 17, 32, 66, and 84, respectively. In fact, two of the most frequently used terms associated with the mega-category Theory and Methods were disciplines themselves, i.e. information science and library science and computer science. All four periods had a large spike in the number of uses of information science and library science The sudden increase of term changes and sporadic use of certain subject terms demonstrated that the Theory and Methods mega-category’s association with institutional repositories was almost non-existent for a time, but later had a rapid increase. For example, international, open access, and open archives initiative (OAI) were subject terms that had the largest spike in Period IV. Interestingly, consistent terms that appeared in the megacategory Theory and Methods were information science and library science and computer science.

There were also fewer terms that were being used in Theory and Methods after Period II that had sudden spikes; this finding demonstrates that more terminology that is consistent was being used in institutional repository research. Therefore, it is an indicator that the field of institutional repositories was becoming more consistent.

Application and management

An overwhelming amount of terms in the mega-category Application and Management (Fig. 7) had an overwhelming number of terms that emerged quickly and rapidly disappeared as evident by the large spikes and drops in numbers. Terms like data repository, digital repository, discipline, institutional repository, and open access all provide more gravity to the lack of consistency in regards to terminology and definitions of institutional repositories, as many of the definitions of the field had been changing as more research was conducted, making it more diversified. New applications and management styles were beginning to emerge within institutional repositories as a whole. Consequently, it was an appropriate time to begin the transformation into application and management as Periods II and III respectively covered years 2002–2009. In conjunction, the early 2000s brought a vast assortment of new and improved technology applications that had not existed in the 1990s; this finding is further strengthened by the data indicating the term institutional repository was the most used term for the mega-category. In addition, the number of terms used in each period increased from an average of appearance from 1.11, 2.38, 4.6, 4.25 in each period respectively. Implementation and technique Interestingly, Implementation and Technique was the only mega-category that did not have any consistent terms meaning terms that changed every other period. Many terms appeared consistently throughout all four periods, but in lower numbers. Instead, the majority of terms that had large spikes appeared at the end of Periods III and IV. For example, information system and information retrieval were two terms that had the largest spikes in the mega-category Implementation and Technique with an average of 10 appearances per term instead of 1.65. This finding is an indicator that a lot of development was spent on institutional repositories over the later periods. It also demonstrates that it was not until recently that studies were published in regards to implementation and technique, which was evidenced by the large influx of case studies that were published in the mid-2000s; this finding was strengthened by the data that indicates that information system was the most highly used term in the mega-category. Figure 8 demonstrates the increasing amount of consistent research being conducted across the four time periods. Terms like experimentation, resource sharing, and web technology rapidly decreased in use in Period IV, which further exemplifies the leveling out of experimental design of institutional repositories; this point is further confirmed by the MDS results found in Period IV. For example, the three clusters identified in Period IV were publication, management, and institutional repository. Terms like continuum, open system, and users of information retrieval all allude to the next phase of development in the research and evaluation. Period IV also had the largest number of subject terms associated with it; this could be an indication that institutional repository studies have been becoming more diversified, and the field is expanding.

Summary

Important themes and patterns could be formed from the parallel coordinate analysis. Observations from the parallel coordinate analysis revealed three major themes. One major theme was the maturity of institutional repositories as a field over time, the second was the fluctuation and developmental status of institutional repositories until Period IV, and the third theme was the emergence of the discipline information science and library as the strong generator of institutional repository research. Overall, Period III demonstrated the most dramatic changes in research in all three of the mega-categories. Maturity and development of the field was evidenced in the mega-category, Theory and Methods. Term categories were the lowest in the mega-category Theory and Methods; this finding demonstrates that terminology that is more consistent was used in institutional repository research. Therefore, it is an indicator that the field was becoming more consistent. In addition, patterns illustrated in Implementation and Technique showed the most noticeable changes in subject term use as far as large spikes of term usage and periods of equalization. The Application and Management mega-category had a lack of consistency in regards to terminology and definitions of institutional repositories, as many of the definitions of the field had been changing as more research was conducted, and more information continues to be gathered. An analysis of the top-five subject terms revealed relationships about the developing research areas of institutional repositories and the field’s future prospects, which is demonstrated in Table 10. In addition, Table 10 further extrapolates how and when the terms are focused into specific subjects, making it easier to visualize how various aspects within institutional repositories have developed. Information science and library science was the top subject term used in the mega-category Theory and Methods. The megacategory, Application and Management had institutional repository as the top term used, and mega-category, Implementation and Technique had the term information system. The pattern illustrated in this study from the parallel coordinate analysis is that the discipline information science and library science has emerged as the primary discipline conducting research about institutional repositories. The term institutional repository also had the highest concentration of term usage out of similar formally used subject terms, for example, digital repository. Finally, information systems are being discussed in terms of implementation and technique within institutional repositories, thus, preparing for the continued use and change.

Discussion

The results reveal a great deal of information regarding research on institutional repositories. The acceptance of the hypotheses H10 and H20 means there were no significant differences between subject terms within the mega-categories, Theory and Methods and Application and Management, across the four time periods, respectively. The acceptance only suggests that the numbers of unique subject terms in each of the mega-categories (Theory and Methods, and Application and Management) stayed relatively stable over the four periods. However, the actual terms within a category might change greatly. For instance, the subject term overlapping rate in the category Theory and Methods for the four periods was only 17.7 %, while the subject-overlapping rate in the mega-category Application and Management for the four periods was only 31 %. A low term overlapping rate indicates a large change of terms in a category. For example, the mega-category, Application and Management, had only 2 consistent terms that appeared throughout all for periods, but had 28 terms that appeared and disappeared. In addition, the mega-category Theory and Methods saw the term overlapping rates for adjacent Periods I and II, II and III, and III and IV to be 5, 5, and 3.7 % respectively. The mega-category, Application and Management had the overlapping term rates for adjacent Periods I and II, II and III, and III and IV to be 12, 5.75, and 3.5 %, respectively. Analyses show that although the Chi Square analysis results do imply the total numbers of subject terms across the four periods stayed relatively stable, the term overlapping percentages imply that a significant amount of term change was occurring within the mega-categories Theory and Methods and Application and Management. The findings of the Chi Square inferential analysis further emphasize the point that a significant portion of foundational research took place in Period I, but not in the later periods. In addition, the rejection of the hypothesis H30 implies that the mega-category Implementation and Technique was the most active category in terms of research topics on institutional repository. The findings were echoed in the new merging term category. The subject term overlapping rate in the mega-category Implementation and Technique for the four periods was 42 %. The mega-category Implementation and Technique had low overlapping term rates for Periods I and II, II and III, and III and IV (1.5, 7, and 13.7 %, respectively). It means that in the mega-category Implementation and Technique for the four periods the significant changes were reflected in not only the number of subject terms, but also the contents of the subject terms. However, in order to more thoroughly analyze the changing intricacies of institutional repositories, the MDS analysis and parallel coordinate analysis were necessary to dive deep into the subject matter. For example, much of the research in Period I was theoretical with many references to database systems. The emergence of these five clusters in Period I indicated that it had a major focus on general development of institutional repositories. Period I was also the longest of all four periods indicated in this study, but the justification was that it was not until Period II that more studies were being produced, as more case studies of institutional repositories were emerging in the early 2000s. Other interesting observations in Period II included changing definitions of institutional repositories, as well as how to discover whether an institutional repository was successful. A significant portion of research also focused on future ideas and development and the developmental process for institutional repositories; this dynamic became very prominent in the last year (2005) of Period II. Periods I and II brought a number of different ideas and possibilities for institutional repositories; Period III took those foundational ideas from Periods I and II and steered research questions outward toward a variety of new resources and users. As a result, there was a significant amount of material that was available during this period. The case studies were quite exhaustive, but there were not extensive evaluation or summarization studies conducted. As a result of the foundational case study basics illustrated in Period III, more case studies were conducted in Period IV than in any other periods. The case studies conducted in Period IV have a strong focus on end users. Recommendations were also beginning to be made in Period IV. For example, Asunka et al. (2011) were specifically focused on the possible integration of social networking tools to aid in communication among faculty involved with institutional repositories. Going back to Graham (1995), when institutional repositories were first beginning comprehensive development, Graham stated that librarians (and information professionals] should be treating digital research libraries or institutional repositories in the same manner as adding material and collections to the physical library. The same approach should be conducted for institutional repositories. The breadth and depth to which studies were covered in Period IV provides the most complex and dynamic clustering of the four periods covered in this study. Again, this point is strengthened by the acceptance of H10 and H20 from the Chi Square inferential method. The Chi Square analysis findings observed there were no significant differences in the number of between numbers of subject terms with the mega-categories Theory and Methods and Application and Management across the four time periods, respectively; however, the terms changed greatly within those mega-categories. For instance, the subject- overlapping rate in the mega-category Theory and Methods was 17.7 %, while the subject overlapping rate in the mega-category Application and Management was 31 %, indicating there might be a number of developments occurring during the early stages of institutional repository foundational periods. The MDS analysis also addressed the related disciplines associated with institutional repositories. Early on in Period I, clusters observed yielded a finding that further illustrated the newly developing nature of institutional repositories. The high use of subject terms such as information science and library science, institutional repository, scholarly communication, and metadata, indicated that information science and library science, rather than computer science, would be the discipline to forge the growing areas of institutional repositories. The results from the parallel coordinate analysis revealed noticeable characteristics about institutional repositories. The majority of these results are complimentary to the MDS findings. For example, there were few subject terms associated with the megacategory, Theory and Methods, until Period III, which saw a large spike, which continued through Period IV. The mega-category Implementation and Technique started in the middle, as compared to the other two mega-categories. This mega-category also had a noticeable spike in Period III, but then saw a decrease moving into Period IV. The megacategory Application and Management began in Period I as the largest period in terms of subject terms associated with it. The category gradually began to decrease in subject terms moving forward after Period II. The uniqueness of this study lies in the longitudinal nature and the use of information visualization, MDS, and parallel coordinate analysis to provide information regarding the field of institutional repositories. The findings were further strengthened by the hypotheses tested from the Chi Square inferential method. A combination of research methods provides a portal to the research being conducted, and analyzed the research in a way that a number of themes and patterns were able to be discovered and observed. The observations from the parallel coordinate analysis revealed three major themes: the maturity of institutional repositories as a field over time, the fluctuation and developmental status of institutional repositories until Period IV, and the emergence of the discipline information science and library science as the strong generator of institutional repository research. These observations provide insight into the possibility of the potential guidelines for institutional repositories. In addition, this study used subject terms most commonly associated with institutional repositories across the four time periods. The knowledge gained from the terminology usage in each period provides more insight into potential revision of the thesauri of terms and subject directories related institutional repositories, which may be helpful for future studies. Through the identification of frequently used subject terms in the findings, information has been gained regarding the history, development, and future studies within institutional repositories.

Limitations

Major research on the topic in English was published in WoS, but it is recognized that other research on the same topic was published outside WoS like conference proceedings, books, and non-English speaking journals was not included in this study. As a result, this dataset is biased towards United States and Europe and English speaking countries. There is a great deal of future research that needs to be conducted.

Conclusion

Institutional repositories enable institutions to share and provide scholarly access and information to a wide audience. Through collaborative partnerships, institutions can create institutional repositories that preserve a wide array of material for generations. More and more institutions are discovering the importance of building relationships with one another to work together to develop institutions that ensure the preservation and availability of their material. Through continued collaboration and evaluation of digital repositories, a highquality resource can be preserved. The creation of an institutional repository provides an extremely rich resource for the public to utilize in many new ways. However, the development of institutional repositories occurred very organically. As a result, the field is still moving into evaluations and potential guidelines. The primary purpose of this study was to have a better understanding of the field of institutional repositories. The end goals of the study were to provide useful information to researchers and practitioners to help develop a scheme to organize information and enrich thesaurus and subject directories. Through the identification of frequently used subject terms in the findings, information can be gained about the history, development, and future studies within institutional repositories. Using MDS, visualization, and parallel coordinate analysis, an in-depth account on the intricacies of institutional repositories was found. The findings of the Chi Square inferential analysis strengthened the MDS and Parallel Coordinate analysis by the acceptance of the H10 and H20, and rejection of the H30; thus meaning, there were no significant changes that took place across the four time periods for the number of subject terms associated with the mega-categories, Theory and Methods and Application and Management, but were significantly different for mega-category Implementation and Technique. In addition, the MDS and parallel coordinate analyses were able to extract specific relationships among the subject terms, and provided an additional way to view the development of institutional repository research. The unique research methodology of this paper demonstrates a new way to pull together various methods of analysis and draw conclusions from a developing field. An analysis of the research indicates that, as a whole, the development of institutional repositories has been a long, interesting development. The MDS and parallel coordinate analyses in conjunction with temporal analysis revealed that institutional repositories transitioned through several phases of development. The development of institutional repositories began with researchers from within the computer science field, which took a theoretical and technical standpoint. Then, practitioners and researches within information science and library science became more involved with the development and application of institutional repositories. As a result, more academic libraries began to explore different possibilities for development and management. In 2006, institutional repository research brought many new case studies, which included new information regarding metadata, organization, and technology methods to create institutional repositories. The purpose of the temporal analysis was to gather institutional repository data to provide a rich array of material to discover developments and changes in the field over time. Consequently, this study made a clear division of four time periods. The MDS analysis found information science and library science to be the discipline that will continue to work strongly with institutional repositories to develop and conduct further research, rather than computer science. In addition, the parallel coordinate analysis expands upon the MDS findings. If the MDS analyses focused on each of the periods, then the parallel coordinate analyses concentrated on the connection between the periods. Through the categorization of the subject terms across the four time periods, a deeper understanding of the developments of institutional repositories could be made. The important themes and patterns can be formed from the parallel coordinate analysis. Observations from the parallel coordinate analysis revealed three major themes: the maturity of institutional repositories as a field over time, the fluctuation and developmental status of institutional repositories until Period IV, and the emergence of the discipline information science and library science as the strongest generator of institutional repository research, which further emphasizes the MDS findings. This study does not come without limitations. As with the nature of visualization studies, a certain amount of interpretation was left to the researchers when categorizing the data. In addition, despite the breadth to which WoS gathers a significant amount of research within different disciplines, it is by no means able to gather all of the studies ever published on institutional repositories, especially those published on the Internet. Many questions are still unanswered in regards to institutional repositories. Fortunately, more institutions are discovering the importance of building relationships with one another to work together toward developing institutional repositories. Further insight into a continuation of usability studies from a log analysis and qualitative user analysis will be needed in order to ensure the best institutional repositories design, which will increase functionality and preservation features for generations.

E-research and Data

Technology has been forming and transforming the nature and scale of e-research. The advance of Web technology opens up new ways for researchers to communicate. The capability to collaborate online encourages cross-discipline research.Researchers from different parts of the world, carrying different expertise, readily get in touch with each other and engage in online discussion and exchange. Theyshare files in a wide variety of formats, containing both text-based and non-text-based research

Objective

In a 2008 document that provides guidance for building policy framework, RIN states the objective of data management as ensuring that “ideas and knowledgederived from publicly funded research should be made available and accessible for public use, interrogation, and scrutiny, as widely, rapidly and effectively aspracticable.” The document sets out five principles as a starting point for drawing up policy framework, with responsibilities shared among all stakeholders:universities, research

Institutional Repositories for Research Data

The present picture of research data management can be described as patchy. Although some research communities have established norms and standards fordepositing data, with well-structured data repositories in place, the reality remains that the majority of the data in most disciplines are left unattended after thepublication value is extracted, regardless of whether the data are in digital form or not. Data from research projects in universities, being an essential component of research

Data Archiving Experiences and Efforts at Universities

Data Archiving Experiences and Efforts at Universities Many universities are implementing data archiving at their institutional repositories with different approaches and levels of ambition. Purdue University adopted a “distributed approach.” E-Data , the designated repository for research datasets built with Fedora, is under development by the Distributed Data Curation Center (D2C2). It serves as a platform for experimentation in data curation. Disciplinary librarians had the responsibility to solicitresearch datasets and used a list of standard

The HKUST Institutional Repository

The HKUST Institutional Repository (IR) was launched in February 2003. Powered by DSpace, the IR was built with the aim to create a permanent record of thescholarly output of HKUST in digital format and to make the Repository globally and openly accessible. In other words, the defining characteristics of the HKUSTIR are

• being fully open access;

• focusing on scholarly output. Other community output such as students work and teaching objects are not included.

The Case Study Sample

The author conducted a case study to survey the common treatment of research data at institutional repositories. The study aims to find out:

• What kinds of data are currently being archived at institutional repositories

• How data are accommodated in different repository software, especially in a DSpace environment

• What measures are effective in making the data discoverable, accessible and usable. As a starting point, the author used OpenDOAR, a major directory of open access repositories. In February

Repository Terms

For effective discussion, the author uses the terminology of DSpace when referring to repository structure and components:

• Organization structure—a repository is divided into communities ; each community contains collections , which are groupings of relatedcontent.

• Items —the basic unit of a repository. An item consists of a set of metadata and bundles of bitstreams.

• Bitstreams —computer files deposited in a repository. Each bitstream is associated with one Bitstream Format for interpretation.

Data Types

To set

Descriptive Information

Datasets are much more versatile than text-based research output. Adequate description is crucial for data management to facilitate data discovery, access and use.Productive descriptive information for data should cover:

• the origin—the context in which the data were generated or gathered, and the methods

• the “table of content”—the list of bitstreams, and what they contain

• the use instruction—how to approach the datasets, what software is required

• use policy—any use or license agreement.

Usability Issues

Even when a dataset is found and accessed, it has to be readable and interpretable in order to be usable. A few issues need to be considered for satisfactory datausability.

Datasets and Related Items

Datasets seldom exist as standalone research output. To facilitate proper use, a dataset should be clearly linked to its related files, such as derived data, researchpapers resulting from the data and previous or later version of the dataset. In DSpace, related items can be presented in a number of ways. One or more descriptionfields may be employed to store the information, either in plain text or hyperlinks to other items in and beyond a repository.

The observation from the sample indicates

Way Forward for the HKUST Institutional Repository

With current institutional culture and human resources available to the Library, the pilot for hosting data at the HKUST IR would be most appropriate to adopt abottom-up approach. Datasets would be ingested into the existing IR, with an additional set of policies and measures to address their special requirements.

The Next Steps

Data management is not a library-specific issue. Libraries may initiate projects as start-up attempts to address the issue, but the ultimate success depends onsupport from other stakeholders in the scholarly communication cycle. Efforts and projects that aim to elevate awareness of research data management amongresearchers, faculty, university administration and funding agencies would be conducive. Mutual understanding is the core of successful collaboration. Although it would be instrumental

Conclusion

Research data management has become a pressing issue. While the need for effective data management is rising everyday with the incessant production of data inall research fields, a myriad of issues are yet to be resolved before a satisfactory system comes into being. The hurdles, however, have not deterred institutions andinterested agencies from developing different systems and models that attempt to address the data issue. Universities that are operating institutional repositoriesare

Acknowledgment The author thanks Dr. Samson Soong, Ms. Corrie Marsh, and Mr. KT Lam for their encouragement and advice.

Academic institutions often have financial interests related to research conducted on campus, such as stock or equity in private research sponsors, intellectual property rights, or relationships with commercial donors (Association of American Medical Colleges 2002; Association of American Medical Colleges and Association of American Universities 2008; Institute of Medicine 2009; Resnik and Shamoo 2002). A conflict of interest (COI) for an academic institution can be defined as a situation in which the institution or its leaders (such as presidents, chancellors, vice presidents, deans, or department heads) have interests that may compromise judgment or decision-making concerning its primary professional, ethical, or legal obligations or academic aims (Institute of Medicine 2009; Resnik and Shamoo 2002). Institutional COIs can create ethical problems by threatening the objectivity and integrity of research and public trust in the institution, investigators, or the research enterprise (Resnik 2007, 2011). In one of the most well-known cases involving questions about the impact of financial interests on research (the death of Jesse Gelsinger, a research subject/participant in a Phase I gene therapy trial at the University of Pennsylvania in 1999) the principal investigator, James Wilson, and the university had significant financial interests at stake. Wilson and the University of Pennsylvania both held a substantial amount of stock in Genovo, a company that sponsored gene therapy research on campus, and they both owned patents on gene therapy methods (Resnik 2007). Much of public scrutiny following Gelsinger’s death focused on the financial interests related to the trial and how they may have affected the conduct of the study or its oversight (Association of American Medical Colleges 2002). Gelsinger’s family sued Wilson and the university, alleging that Gelsinger was not adequately informed of the risks of the research as well as the financial interests of the investigator and institution (Resnik 2007). Some other hypothetical (though realistic) examples of institutional COIs include the following:

• In the clinical trial of a medical device sponsored by a private company, the vice president for research and dean of the medical school write a letter endorsing the study to the committee that oversees human subjects research and ask to attend the meeting in which the study is reviewed. The university and several investigators own stock in the company. Members of the human subjects committee fear that their careers could be harmed if they do not approve the study, or if they require changes that could affect the conduct of the research.

• A university accepts a $300 million gift from a wealthy alumnus who is the chief executive officer of a contract research organization he formed 15 years previously. According to conditions set forth in the gift, the university’s school of public health will be renamed after the donor, and the curriculum in the school of public health will include several courses on clinical trial design, management, and regulation. There will also be an endowed professorship in clinical trial design, management, and regulation. Faculty members at the school of public health are concerned because they feel that the donor is inappropriately influencing the curriculum.

• A university enters into a research and development partnership with a robotics company. The company will sponsor research through a new robotics center on campus named after the company. The company will provide funding for a new building to house its robotics research, equipment, and several faculty positions. Under the terms of the arrangement, faculty whose research is sponsored by the company cannot publish results or share data without the company’s permission. Research taking place in the new building will be kept strictly confidential, and access to the building will be restricted. University faculty members are concerned that this collaboration undermines academic norms by interfering with open communication and publication.

• A prominent biomedical researcher at a university medical center who holds several patents and brings in millions of dollars per year in grants is accused of data fabrication and falsification by a junior faculty colleague. The dean of the medical school and an institutional official in charge of reviewing the allegation decide not to pursue it in order to protect the researcher and the institution’s financial interests. When the junior faculty member accuses the dean and institutional official of covering up the problem, her employment is terminated.

This paper will examine institutional financial interests related to research in greater depth and provide guidance for academic institutions.

Institutional Conflicts of Interest

Most of the scholarship, research, and policy analysis on COIs in research has focused on individual financial interests. There is a sizeable body of evidence demonstrating that sources of funding and investigator’s financial interests can influence research outcomes (Friedberg et al. 1999; Krimsky 2003; Lexchin et al. 2003; Resnik 2007; Ridker and Torres 2006; Sismondo 2008; Stelfox et al. 1998). A leading explanation of the relationship between financial interests and research outcomes is that investigators or companies with financial interests at stake may make decisions that tend to bias research in favor of those interests (Lexchin 2012; Resnik 2007). For example, companies can affect research outcomes by funding studies they expect will produce results favorable to their interests, and not funding or de-funding those they do not expect to yield favorable results. If the company is the dominant funder of research related to a particular topic, the published literature will tend to reflect the company’s research agenda (Michaels 2008). Investigators can bias research by using a study design that is more likely to support their hypothesis than other designs or by analyzing or interpreting data in a way that favors their interests or the interests of the sponsor (Michaels 2008). Companies may decide not to publish data that are unfavorable to their products (Lexchin 2012). In extreme cases, investigators or companies may fabricate or falsify data in order to produce results favorable to their interests (Krimsky 2007). While the potential effects of financial interests on science demand careful scrutiny, it is important to realize that financial interests are only a risk factor for bias; they do not automatically invalidate research (Shamoo and Resnik 2015). Even so, it is important to acknowledge and address financial interests in research to reduce their impact (Resnik 2007). One of the main ways that institutional financial interests can affect research is by compromising independent review and oversight of research conducted at the institution. Independent review and oversight is import for ensuring that research meets scientific, ethical, and legal standards. For example, if an institution or its leaders have a significant financial stake in a clinical study conducted on campus (such as the Gelsinger study mentioned above) then committees that oversee the research may feel some pressure to approve the research or refrain from criticizing it (Resnik and Shamoo 2002). If a well-funded investigator is accused of research misconduct, institutional leaders may be reluctant to conduct a thorough investigation of the allegation because they do not want to jeopardize millions of dollars of research support or face adverse publicity (Resnik 2008; Rivlin 2004; Shamoo and Resnik 2015; Smith 2006). Because institutional COIs can compromise review and oversight, institutional COIs may affect more people than individual ones. For example, dozens of people could be affected by a research and development partnership with a private company, a corporate gift with strings attached, or lax oversight of a privately-funded clinical trial (Resnik and Shamoo 2002).

Dealing with Institutional Conflicts of Interest in Academic Research

While there are some well-recognized best practices for dealing with individual COIs, such as disclosure to independent parties, conflict management, or prohibition in some cases, there is no consensus regarding how best to deal with institutional COIs (Shamoo and Resnik 2015). Indeed, a recent study by the author and colleagues (Resnik et al. 2015) found that only 28 % of 100 of the top U.S. academic research institutions, ranked by total research funding, had an institutional COI policy. The study also found that having an institutional COI policy was positively associated with total research funding and that there was considerable variation in the content of the policies with regard to committee structure and reporting. The main reason why there has been a lack of policy development in this area is that federal granting agencies and journals have not adopted regulations or guidelines for dealing with institutional COIs in research. U.S. research institutions have developed COI policies for investigators because this is a requirement for receiving funding from the National Institutes of Health (NIH) or National Science Foundation (NSF) (Resnik 2007). The NIH and NSF rules require investigators who receive grants or contracts to disclose significant financial interests to the institution and the agency (National Institutes of Health 2013; National Science Foundation 2005; Public Health Service 2011). Although a report by the Department of Health and Human Services’ Office of Inspector General and the Institute of Medicine (2009) urged the NIH to require institutional grant or contract recipients to develop policies pertaining to institutional COIs, so far the agency has not acted on this recommendation (National Institutes of Health 2013; Wadman 2011). At the same time, the NSF COI policy does require grant or contract recipients to address institutional COIs (National Science Foundation 2005). Most scientific journal policies require authors to disclose their own COIs and sources of funding, but most do not address institutional COIs (Cooper et al. 2006). Dealing with institutional COIs effectively can be more difficult than dealing with individual ones for at least two reasons. First, since the institution may have many different interests and departments, schools, or committees that deal with those interests, it may be difficult to recognize institutional COIs (Institute of Medicine 2009). No single person, office, or department may be aware of all of the institution’s conflicting interests. Second, to manage COIs effectively, it is essential to make use of an independent party who does not have a conflicting interest but who has authority to take actions to deal with the COI. However, it may be difficult to identify independent parties at an institution who can manage its COIs, since people in leadership positions will often have interests that are directly related to the institution’s interests or they may not have the authority to take actions to deal with its COIs (Institute of Medicine 2009). A dean, vice president, or president may not want to raise concerns about a grant, contract, or industry collaboration that brings considerable money or prestige to the institution. The close connection between the interests of the institution and its leaders may even create a climate of corruption in which leaders interfere with institutional oversight and decision-making or turn a blind eye toward ethical, legal, or other problems (Shamoo and Resnik 2015). Although some of the most egregious cases of corruption have involved lax oversight of athletic programs at universities (such as a cover up of sexual misconduct involving a football coach at Pennsylvania State University), corruption involving academic research also occurs (Institute of Medicine 2009; Wolverton 2012). The most important resource that academic institutions have for dealing with institutional COIs is a board of trustees or similar body. The board of trustees consists of members of the larger community, such as business leaders or alumni, who are not employed by the institution. Many boards also include the student body president and other ex-officio members from the university community. Members of the board are responsible for protecting the institution’s integrity, reputation, and financial solvency and promoting excellence in education and research. Boards are supposed to exercise independent judgment in managing the institution. Though the board of trustees delegates decision-making responsibilities to different officials and organizations, it has final authority over all institutional decisions and there are numerous recent legal cases where trustees have been found liable for failing to adequately undertake their fiduciary responsibilities (Institute of Medicine 2009). The board of trustees could appoint a committee composed of board members as well as other independent parties with expertise to deal with institutional COIs. Alternatively, the board could appoint a committee that reports directly to the board (Duke University 2009). The trustee committee could receive reports of institutional COIs and decide whether they should be disclosed to appropriate parties (such as funding agencies, journals, or research subjects), prohibited, or managed through additional oversight (Institute of Medicine 2009). The trustee committee could also receive reports from different parts of the institution that may have information about COIs (Resnik et al. 2015). For example, most universities have formed committees to review faculty COI disclosures (Institute of Medicine 2009). These same committees could determine whether institutional interests are also affected by the situations that create faculty COIs and whether they should be reported to the trustee committee. Administrators working in various offices and departments could also report institutional COIs to the trustee committee. Some universities use the committee that oversees faculty COIs to deal with institutional COIs (Resnik et al. 2015; University of North Carolina at Chapel Hill 2009). Others appoint a special institutional COI committee (Resnik et al. 2015; Washington University 2009). While these COI committees may have the expertise to deal with institutional COIs, they may not have sufficient independence or authority to affect decisions made by the academic institution. For example, such committees might not be able to prevent the institution from entering into questionable relationships. A committee at the level of the board of trustees would have the requisite independence and authority to deal with institutional COIs (Institute of Medicine 2009). Lack of awareness may be a significant obstacle to forming an institutional COI committee. Trustees may not be familiar with the potential impact of institutional COIs on research and education, or the need to manage these situations objectively. This problem can be overcome if institutional leaders educate trustees about institutional COIs and how they may impact the university. However, some of the trustees may have financial interests of their own that are related to the institutional COIs. For example, a trustee could own stock in a company that sponsors research on campus. One study found that one-fourth of private colleges or universities have financial ties with trustee-affiliated companies (Fain et al. 2010). While this is an important concern, it can be addressed if trustees are aware of COI issues, as well as their own COIs. Thus, institutional COI policies should extend to trustees. In some cases, it may be necessary for trustees to divest themselves of the interests that create COIs. Fortunately, many academic institutions have begun to address the problem of trustee COIs. The Association of Governing Boards of Colleges and Universities has adopted a COI policy that provides useful guidance for board members (Association of Governing Boards of Colleges and Universities 2009). The policy states that board members should disclose all COIs related to institutional business and should recuse themselves from decisions pertaining to business matters in which they have a COI, unless the board determines that their participation in such decisions offers compelling benefits to the institution. Academic institutions can also develop policies which prevent communication and influence between different administrative offices and functions, such as grants and contracts, compliance and oversight, human resources, legal counsel, and fundraising. This separation can help to protect the integrity and independence of university decisions by preventing inappropriate interference in decision-making. Many of the concerns about institutional COIs pertain to concerns about inappropriate influences on decisions, such as the possibility that fundraising needs may shape curriculum development, or that contracts with companies may affect publication of results. Though many universities and colleges already have such policies in place, those that do not can minimize the impact of institutional COIs by ensuring that different functions remain separate (Resnik and Shamoo 2002). Academic institutions can also develop policies to ensure that committees and offices that deal with different functions, such as human subjects research, animal research, biosafety, regulatory compliance, equal employment opportunity, and curriculum development, are able to make independent decisions without outside pressure or fear of reprisal. For example, committee meetings can be kept confidential, with only committee members or those invited by the committee in attendance. Meeting minutes should summarize the discussion without identifying individuals. Committee members should have no financial or other interests related to business conducted by the committee (Resnik 2008). Academic institutions can also establish private research foundations to own stock and intellectual property, engage in fundraising activities, and provide venture capital for start-up companies. Locating these activities in a private foundation can soften the impact of institutional financial interests even though it does not eliminate them. That is, COIs would remain because the private foundation would be closely aligned with the university and may have an interlocking board of trustees, and leaders in both organizations would probably know and influence each other (Resnik and Shamoo 2002). However, establishing research foundations can help protect university or college decisions from the direct influence of institutional financial interests and promote public trust (Moses and Martin 2001). Finally, academic institutions can support educational activities, such as seminars, lectures, and online learning modules, concerning institutional COIs. These educational activities could supplement education that already occurs on individual COIs, and would help to raise awareness of the issues among faculty, administrators, staff, and students. Educational activities should provide learners with information about institutional policies on COI disclosure and management.

Conclusion

Financial relationships in academic research can create institutional COIs because the financial interests of the institution as well as those of institutional officials may inappropriately influence oversight and decision-making. Though institutional COIs can be difficult to deal with, universities and colleges have some resources and policy options at their disposal, such as establishing COI committees that involve the board of trustees in conflict review and management, developing policies that shield institutional decisions from inappropriate influences, and establishing private foundations to own stock and intellectual property, and to provide capital to start-up companies. Although funding agencies do not currently require institutional grant or contract recipients to address their own COIs, universities and colleges should develop institutional COI policies themselves without such prodding, in order to protect the objectivity and integrity of academic research and promote public trust. Acknowledgments This article is the work product of an employee or group of employees of the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH). However, the statements, opinions or conclusions contained therein do not necessarily represent the statements, opinions or conclusions of the NIEHS, NIH, or the United States government.

As more nursing research is conducted, there is and will continue to be increased interaction between the nurse researcher and the institutional review board (IRB). Research involving human subjects must be approved by a research committee. Many healthcare institutions and universities have their own IRBs. However, others may contract with an outside agency. Although various IRBs may operate differently from institution to institution, they all must follow the same federal guidelines and are highly structured. 1 Anyone who conducts research or is involved with clinical research must be familiar with the working of the IRB.2 The purpose of this article is to introduce the role of the IRB to the critical care bedside nurse. Institutional review boards were formed as one result of the Nuremburg Code, which provides a basis for all ethical guidelines regarding research using human subjects.3 The code was developed after World War II in response to the horrific war crimes. In 1953, the National Institutes of Health in the United States implemented a mechanism to protect human subjects. This was the beginning of the modern-day IRB. Unethical research has been conducted by many individuals and countries. The war crimes of World War II come to mind first, but the United States is not without blame concerning unethical research. The Tuskegee Syphilis Study was conducted in the United States in 1932 when black men with syphilis were not treated even though treatment with penicillin was available.2 However, since the formation of the Food and Drug Administration (FDA) in the early 1960s, many efforts have been made to protect the rights of humans in clinical trials. The protection of human rights is paramount in all research. The United States now requires that any research involving human subjects that receives federal funding be approved by an IRB. In addition, healthcare facilities and universities may also require approval from an IRB for any research conducted in their respective facility before any data can be collected. Also, most scholarly journals also require evidence of IRB approval before any results will be published.

PURPOSE OF THE IRB

The primary purpose of the IRB is to ensure that the rights of human subjects are protected before, during, and after their participation in clinical trials. The IRB ascertains that subjects have been given adequate information about the research and have voluntarily given informed consent. An IRB functions under the guidelines and regulation of the FDA.2 The IRB must take measures to review research proposals to determine that ethical guidelines are followed. These measures are (1) determining if the research is reasonable, (2) ascertaining that all ethical guidelines have been followed, (3) acting as a layer of protection for the subjects and the institution, (4) reviewing all research protocols before they are implemented in the institution, (5) continually taking effects to protect research subjects, (6) determining if there is any potential harm to the subject and then taking measures to protect subjects from harm, (7) determining if the potential risk to the subject is reasonable, (8) reviewing and approving informed consent documents, (9) continuing to review the progress of the research while the study is underway at least annually; and (10) monitoring for any adverse events.

IRB MEMBERSHIP

An IRB is a group of people who review research proposals in an effort to protect the rights of human subjects. The IRB must have at least 5 members of varying backgrounds including scientists, nonscientists, representatives of the community, representatives of the institution, and at least 1 person who is not affiliated with the institution.1-3 Potential members include nurses, physicians, pharmacists, ethicists, clergy, attorneys, and community advocates. Efforts must be made to have a diverse group of members.4 The IRB must have at least 5 members of varying backgrounds. LEVELS OF REVIEW There are 3 levels of review for a proposal when it is submitted to the IRB for approval. These are (1) exempt, (2) expedited, and (3) full.2,3 These levels Bdetermine the type of oversight the IRB requires over the course of the project.[2(p293) The chairperson of the IRB determines the level of review.1-3 Exempt review is one that does not necessitate regular IRB review. Research that meets the criteria for exempt review is that which does not involve invasive procedures and requires examination or collection of existing data such as chart review. The subjects’ anonymity must be maintained, and there must be no potential foreseeable risk.1-3 Examples of such research may be the completion of a survey, some interviews, and observation studies. Research that poses only minimal risk for the subjects may qualify for expedited review.1-3 In many cases, the Bproposal may be reviewed and approved without a meeting of the IRB.[4(p6) One example of expedited review is the routine review of studies for continued approval that have no major changes to the protocol. Examples of other studies that meet the criteria for expedited review include (1) requirement of an additional amount of blood drawn for other testing, (2) data collection using noninvasive procedures, and (3) research that may be qualitative in nature.1 Many of the proposals submitted for review to the IRB will require a full review. These are the research studies that have more than a minimal risk to the subject. These proposals are reviewed and discussed by all members of the IRB, and measures are taken to ascertain the risks to the subjects as compared with the potential benefits. In addition, the committee will examine the informed consent procedure and will monitor the progress of the research.

THE CRITICAL CARE NURSE’S ROLE WITH THE IRB

Critical care nurses are often asked to help with clinical research. The nurse may be asked to help with the informed consent process, identify potential subjects, and/or collect data. Others will be the primary investigator of clinical nursing research studies. There are some things that all nurses, no matter their involvement in research, must know. First, if a research study is to be conducted, determine if the institution has its own IRB or uses an outside agency. Make certain that the research has been approved by the appropriate IRB before it begins.1 Second, if the nurse wants to conduct a research study, permission must be obtained from the IRB. The researcher must assume responsibility for contacting the IRB and complete the application process for a research proposal.1,2 Also, the nurse researcher must make requested reports to the IRB. Third, all nurses must take measures to protect the rights of patients. The nurse serves as a patient advocate. Make sure that the patient has voluntarily given informed consent to participate in the clinical research and is aware of the potential risks and benefits of the research. Be prepared to answer questions or to refer the patient to the appropriate person.

CONCLUSION

The IRB is a committee designed by federal regulations to protect the rights of human subjects in clinical research. Critical care nurses may be asked to participate in clinical research in a variety of roles and must be aware of the role of the IRB as well as their role in clinical research.

Institutional policies and educational programs related to the use of animals in biomedical research are certainly not restricted to academic medical centers or universities. Indeed, such policies or programs must be integrated into pharmaceutical corporation research and development laboratories and into private or public research institutes where animals are used for biological research. Such policies are also applicable for agricultural or veterinary units, for zoological gardens, and for aquatic theme parks. Well-planned programs and policies are necessary for internal continuity of research programs as well as to inform the lay public about the importance of biomedical research in the pursuit of new therapies and improved surgical procedures. The public must also have an appreciation for aspects of basic biomedical research, and it is incumbent upon the scientist to aid in this understanding. Unfortunately, the lay public or the nonscientist audience may not appreciate the importance of the use of animals in basic biomedical research. In order for the public to better understand why animals are used in biomedical research, they need to first understand some of the basic tenets of science or biology. It has been estimated that only between 5-7% of the U.S. population is scientifically literate (Evans and Frazier, 1993). Further, it is often the nonscience college major that is responsible for the science instruction in the U.S. public schools. About 50% of entering college freshman who major in science change their mind and their major in their first year of studies. It is not surprising that educating the public about the importance of science in biomedical research is rendered difficult because the majority of citizens are not scientifically literate. Despite this lack of understanding, many public opinion polls still reveal overwhelming support for biomedical research even if it involves the use of animals. An institution’s policy, particularly some of the external educational programs, can aid in informing the public about the importance of biomedical research using animals. There must be a consistent balance between internal and external policies. Both internal and external policies may have strong educational components. Presentations or seminars targeted at research into aspects of different diseases is often helpful. News stories of human interest and disease states can often be suitably related to biomedical research findings. Videos and film clips are likewise helpful in informing the public about the importance of animal research. The list below shows a number of factors that should be included in an institution’s external policies as they relate to the use of animals in research: Regulatory and Legal Considerations (e.g., U.S. Department of Agriculture) Membership in national organizations [e.g., National Association for Membership in statehocal organizations Pro-research institutional news and information Community involvement Patient education programs Other (e.g., bioethic coursesflectures, science literacy programs K-12, etc.) It is important for the public to understand that animal experimentation is regulated by stringent policies and guidelines from both the U.S. Department of Agriculture and by the Department of Health and Human Services. Institutional whistle blower policies also contribute to compliance. In addition, vivarium certification by the American Association for the Accreditation of Laboratory Animal Care (AAALAC) and other professional external organizations contribute to strengthening regulatory and even legal consideration. Protocol reviews by peer scientists (IACUCs), serve to foster regulatory and compliance issues as they relate to the humane care and use of animals in biomedical research (Guide for the Care and Use of Laboratory Animals, 1985). It is important for an institution’s external policy on the use of animals in research that the institution communicate with or hold membership in any one of several national organizations that support biomedical research. Certainly, the National Association for Biomedical Research and Research!America understand the national agenda and provide important information to the institution. Institutional memberships in AALAS, Association of American Medical Colleges (AAMC), and other professional scientific and medical societies is recommended. The American Medical Association, American Veterinary Medical Association (AVMA) and the American Dental Association (ADA) all have written policy statements regarding the importance of research and the humane care and use of animals. Such organizations can be either national or local in scope, yet all are beneficial in projecting educational policies to both their constituencies and the lay public. An institution’s external policy necessitates nurturing a proresearch news and media information program. Openness about research projects and research facilities must serve to educate the press or the media about the importance of research. Invitations to the local press should include interviews with the biomedical scientist and tours of laboratories and vivarium, culminating, whenever possible, with an interview with a patient or someone afflicted with a related disease state. The major goal of creating a proresearch media is to correlate the basic research to its potential clinical use or treatment. Sectors of the community can be enlisted to bolster an institution’s external policy. Nonprofit health associations (e.g., American Heart Association, American Cancer Association) are important proresearch partners even though each organization will elect to be focused. Some nonprofit health associations may not choose to participate publicly for fear of losing contributions, but all understand the importance of the use of animals in biomedical research. Community involvement means that local political leaders must also be educated regarding the importance of the use of animals in biomedical research. They must understand the link between basic biomedical research in the quest to understand a particular disease process and the use of local pound animals. Community leaders, both political and nonpolitical, must be proactively educated about the importance of animals in biomedical research. Otherwise, pro-animal rights groups may preempt their understanding with misinformation and sensationalism. Programs targeted toward community involvement should be continuous and not simply a one-time effort. This latter point cannot be emphasized strongly enough. Patient education programs, particularly if the institution’s research programs are conjoined with a hospital, can be an integral component of external policies. Patients who have previously received organ transplants can make particularly good advocates for biomedical research using animals. Patient support groups can likewise play a vital role in this process. Physicians and surgeons can play an important role in patient education programs. Likewise, dentists can relate many oral diseases and procedures to animal research and can inform their patients. If the patient happens to be a pet, such as a cat or dog, then the veterinarian can provide educational materials to the pet owner describing how important animal research is to the well-being of their cat or dog. Finally, external policies can include courses or lectures in bioethics not only to health professionals, but to the lay public. Science literacy programs should be considered an integral part of an institution’s external policies since it focuses upon youngsters at the elementary and secondary level. Most elementary and secondary level science teachers will welcome a partnership with the biomedical scientist. The state junior academy of sciences can be enlisted in these programs. A balanced program involving the use of animals in research also has an important internal component, as outlined in the list below: Administrative support Animal Care and Use Committee Professional veterinarians and support staff Knowledgeable public affairs office (News and Information) Security and legal support Physical facilities of vivarium Educational programs for researchers and nonresearchers Other policies (grants and contracts; occupational medicine program, etc.) Administrative support, both fiscal and nonfiscal is extremely important. The chief executive officer or the university president must be kept informed of the key issues and their implications for the research mission of the institution. Within the sphere of administrative support, there must be coordination and communication between research and development departments, the researcher and the office of public affairs. Research is expensive, and the cost of doing research necessitates strong administrative support through allocation of resources whether it is for maintaining the physical facilities of the vivarium or educational programs espousing the importance of the use of animals in research. IACUCs, like the Institutional Review Boards for the protection of human subjects, are the backbone of the entire internal and external policies pertaining to the use of animals in biomedical research. Not only do such committees provide the guidance and approval for experimental protocols, but they have an important role in regulatory assurance and bioethics. The IACUC must be composed of respected and established scientists, physicians, and veterinarians. It must also have other representative constituencies from the community, the clergy and ideally the institution’s news and information specialist. The IACUC’s effectiveness is only as good as its veterinarian support. The doctors of veterinary medicine provide not only the necessary professional underpinnings to the committee, but they provide the daily monitoring of the IACUC-approved animal protocols. The ofice of public affairs or news and information plays a key role in the success of the institution’s policies-both externally and internally. Community awareness can be provided by the institution’s ofice of public affairs through human interest stories that relate disease states, clinical therapies, and basic research using animals. These links are vital, and the research scientist must participate in order to articulate research findings and their significance to the news and information specialist. Security, particularly within the vivarium itself, but also in individual investigator’s laboratories, must be continuous and vigilant. Limited entry access, proper identification, and surveillance systems are an integral part of security policies. Clear policies on cameras, videos, and visitation to laboratories must be articulated even in public institutions. Since Freedom of Information Act (FOIA) requests are increasingly common, legal advice must be available to the IACUC and to the individual investigator to insure that their rights to privacy are upheld. Clean, well-kept physical facilities not only insure regulatory compliance, but they insure better animal care environments and better scientific outcomes. Healthy animal colonies provide for better scientifically controlled studies, hence resources must be available for the maintenance of animal laboratory facilities. Any institution that has a significant commitment to basic research using animals must have a strong educational and training component. Educational programs may range from communicating with the public about new research findings to media training for research scientists and physicians. Articulate spokespersons, whether former patients or molecular biologists, need to be able to educate not only their peers, but other diverse audiences. In university or medical center settings, the office of grants and contracts or sponsored research is an integral part of any successful program pertaining to the use of animals. Since the grants and contracts office and/or the budget office usually provides the spending authorization for a particular grant, they necessarily are the repository for activation of animal experimental protocols. Likewise, these offices provide the necessary financial information that ultimately inactivates such protocols. Finally, internal policies pertaining to an institution's use of animals requires a program in occupational medicine (Table 1). While such programs vary from institution to institution, they need to include monitoring of tuberculosis among those persons having significant contact with the animals, particularly those laboratories using primates. Serotyping and tetanus and other immunizations appropriate to the person employed in the vivarium must be considered. In summary, policies regarding the use of animals in research must be firmly established and well articulated. The policy must have both internal and external components, and they must be well coordinated within large and complex medical center or industrial research and development laboratories. Education is a key component to these policies since both the lay public and the biomedical community need to work together to increase the understanding and awareness of the value of the use of animals in leading to tomorrow’s new therapies and surgical procedures.

Since the 1940’s and 1950’s, the United States has sought to maintain its competitive edge in science and engineering. Prior to World War II, federal government resources devoted to scientific research and innovation were limited. However, by the end of the war, the United States realized that sustained scientific research could provide new technologies in communications, transportation, and weaponry. One of the best-known examples of the role of science in World War II is the Manhattan Project, which culminated in the creation of the first atomic bomb. In 1957, the Soviet Union launched Sputnik, the first artificial satellite to orbit the earth. Fearing that the Soviets would rise to superiority in math and science, the U.S. federal government further expanded its role of research in science (Lewis, 2005). “Research and development (R&D) is essential for U.S. economic strength, technological leadership, and national security. Strength in science and technology is important because it is an essential ingredient for U.S. economic and military strength” (Lewis, 2005, p. 3). While the federal government has played an active role in funding R&D, its emphasis has been more on development than on basic research which according to Lewis (2005), is critical for the nation’s economy and national security. Basic research has no immediate commercial application or use, but it fuels innovation. Because it takes years to see the results of basic research through products and services, and because businesses and industry seek more immediate return on their investments, private industry is less likely to fund basic research (Lewis, 2005). According to the National Science Foundation (NSF) (2007), colleges and universities spent over $45 billion in R&D expenditures in fiscal year 2005. Of that amount, over $29 billion came from federal sources and over $8 billion came from the institutions themselves. Industry, states/local governments, and other sources funded very little university-based R&D: over $2 billion, almost $3 billion, and over $3 billion respectively. Most federal funds are not applied to the research infrastructure of universities, but rather serves as the main source for specific research projects (Lewis, 2005). When equipment and labs, which build institutional research capacity, are included in a grant, they are usually targeted to a specific project. This leaves institutions to rely more on non-federal sources to increase their research capacity (Marburger, 2006). The U.S. faces a number of challenges as it enters the 21st century: globalization, transition to an information economy, a competitive international business environment, the economic rise of Asia, and sophisticated commercial technologies. Without an increase in scientific research, investment in science and technology, and an enhanced research capacity, the U.S. will not maintain its competitive advantage in national security and economic strength (Lewis, 2005).

A Strategic Plan for Texas

In 1998, Don W. Brown, Texas Higher Education Commissioner, formed a panel of higher education professionals “to review the responsibilities and procedures of the Texas Higher Education Coordinating Board (THECB), and to recommend any changes that would increase its value to Texas higher education’s quality, access, efficiency, and responsiveness to state needs” (Report of the Review Panel, 1998, Background section, para. 1). Commissioner Brown asked the panel to focus on two specific issues facing Texas: (1) how to provide opportunities for individuals to participate and succeed in higher education in Texas, and (2) how to meet the growing demand of higher education over the next 10-15 years while state appropriations of general revenue are not expected to increase at the same rate (Report of the Review Panel, 1998). The panel met with higher education representatives, legislators, the Governor, Texas business leaders who were former higher education institution trustees, THECB members, and Commissioner Brown. The panel also reviewed recommendations from colleges and universities, THECB planning documents, agendas, and legislation outlining the THECB’s responsibilities (Report of the Review Panel, 1998). The Report of the Review Panel, issued in April 1998, made three main recommendations: (1) build on the THECB’s strengths, (2) reassign responsibilities or streamline the procedures for addressing issues, and (3) adopt a long-term view that focuses on the most critical issues facing Texas and creates a public agenda for higher education in Texas. As a result of the initial meetings, a Coordinating Board Planning Committee was appointed and charged with the task of developing a new higher education plan. The new plan included: (1) setting goals, (2) setting dates to reach the goals, and (3) creating a means to measure progress towards the goals. The efforts of more than 1,500 individuals and groups in the higher education community culminated in October 2000 with the THECB’s Closing the Gaps Higher Education Plan (Texas Higher Education Coordinating Board, n.d.a). This document outlines a long-term plan that addresses the four goals of closing the gaps in higher education-participation, success, educational excellence, and funded research over the next 15 years. These four goals were the most critical issues facing Texas higher education from the perspective of the committee and the higher education community. Each goal included interim targets for closing the gaps and success strategies for the state. The purpose of the Closing the Gaps Higher Education Plan was for Texas to develop a higher education system that would keep pace with the continued growth of the state’s economic prosperity. The first goal, participation, addresses the challenge of increasing enrollment in higher education by 500,000 students by 2015. The second goal, success, seeks to increase by 50 percent the number of degrees, certificates, and other identifiable student accomplishments in high quality programs. Excellence, the third goal, seeks to increase the number of nationally recognized programs or services at colleges and universities (Texas Higher Education Coordinating Board [THECB], n.d.b). The fourth goal, funded research, is the issue to be addressed in this study. To meet the fourth goal of funded research, the state initially planned to “increase the level of federal science and engineering research funding to Texas institutions by 50 percent to $1.3 billion” and to “increase research expenditures by Texas public universities and health-related institutions from $1.45 billion to $3 billion by 2015 (approximate 5 percent increase per year)” (THECB, n.d.b, p. 16). In October 2005, this goal was revised to read: “by 2015, increase the level of federal science and engineering research and development obligations to Texas institutions to 6.5 percent of obligations to higher education institutions across the nation” (Texas Higher Education Coordinating Board [THECB], 2006, p. 11). By measuring improved funding to Texas institutions in comparison to other states, the revised goal more clearly expressed the plan’s intent (THECB, 2006).

The Texas Research Development Fund

The Texas Research Development Fund (RDF) and its predecessors, the Texas Excellence Fund (TEF) and the University Research Fund (URF), were developed by the Texas legislature to increase institutional research capacity at 32 participating Texas public universities that do not participate in the state’s Permanent University Fund (PUF). Three institutions that receive funds under the PUF—The University of Texas at Austin, Texas A&M University, and Prairie View A&M University—are exempt from participating in the RDF. The fund links the level of RDF resources awarded to an institution to its success in attracting external funding for research. Specifically, universities participating in the RDF receive distributions based on the average amount of restricted research expenditures per year for the three preceding state fiscal years (Texas Education Code, Chapter 62.095, 2005). Restricted research expenditures are expenditures from a project classified as research, and the funds must be from a restricted funds group (Texas Higher Education Coordinating Board, n.d.c.).

The Study

Thirty-two Texas public universities participate in the Texas Research Development Fund (RDF). However, because one of the 32 universities reports its restricted research expenditures in conjunction with its main campus, for study purposes, there were a total of 31. Table 1 summarizes the 31 Texas public universities that participated in the study, including their Carnegie Classifications. This cross-sectional study employed quantitative methods and focus groups/ interviews. Its purpose was to examine the impact that the RDF may have had in improving the research capacity of participating universities by examining the change in external sources in relationship to the change in the level of RDF resources. Additionally, the study examined the relationship between changes in external resources and selected institutional characteristics. For purposes of the study, research capacity is defined as the amount of restricted research expenditures other than those from the RDF.

The Results

The results of the study indicated that the RDF has made a positive impact on the research capacity of the 31 participating institutions of higher education in Texas. A t-test of repeated measures was performed to determine if the research capacity of the participating Texas public universities had increased significantly since implementation of the RDF. The results indicated a statistically significant increase in the research capacity of the participating institutions of higher education in Texas. Next, an examination was conducted to determine if there was a relationship between the category of institution, based on the Carnegie Classification, and the rate of growth of research capacity. A chi-square test of independence showed no significant relationships between category of institution and the rate of growth of research capacity between fiscal years 2001 and 2005; however, a significant relationship was found between fiscal years 2005 and 2006 and 2006 and 2007. Between fiscal year 2005 and 2006, the phi-coefficient indicated that the Master’s-Large, Master’s-Medium, and Baccalaureate universities had a higher rate of growth in research capacity than the Research-High and Doctoral/Research universities. Between fiscal years 2006 and 2007, the Research-High and Doctoral/Research universities had a higher rate of growth in research capacity than the other lower ranking universities. Each university was categorized as high or low based on the mean rate of change of non-RDF funds for each of the years for which RDF awards were made and sorted into the two groups of institutions mentioned earlier. Figure 1 shows the changes in the mean rate of change in research capacity by year for each of the two categories of institutions.acity is defined as the amount of restricted research expenditures other than those from the RDF.

Lewis (2005) stated that basic research has no immediate commercial application or use, but it fuels innovation. He contends that it takes years to see the results of basic research through products and services. Similarly, Paul Romer (1996), one of the primary developers of the New Growth Theory, characterizes our economy as one in which large upfront costs are incurred; but after the initial work is done, the cost of each additional unit is minimal or nil. Romer says that in a knowledge-based economy, returns increase rather than decrease and that knowledge builds on itself; as society learns more, it gets better at discovering new things (Kurtzman, 1997). Romer suggests that research and development activities, which are associated with long-term economic growth, are the driving force of long-term economic growth (European Commission, 2001). Using this logic and based on the research noted above, the lack of a significant relationship between category of institution and the rate of growth per year of research capacity for the early years of the RDF should not be surprising. The relationships between fiscal years 2005 and 2006 and 2006 and 2007 indicate that significant increases in research capacity are becoming visible four years after the initial institutional investments of RDF (fiscal years 2002 through 2005). Another reason for the lack of a significant relationship between category of institution and the annual rate of research capacity growth in the early years could be due to institutions’ hesitancy in coding restricted research expenditures. This was mentioned by university research representatives in the focus groups/interviews. In other words, institutions may not have accounted for all restricted research expenditures during the initial years of the RDF. In addition, the more experienced and larger universities may know how to increase research capacity consistently through RDF appropriation investments, while the smaller universities are less experienced and knowledgeable in using their RDF funds to increase research capacity. A stepwise multiple regression analysis was conducted to determine which institutional variables among the 31 institutions could be related to the rate of growth of non-RDF funds between the first fiscal year of the RDF (2002) and the latest year available (2007). The rate of growth was the dependent variable, and the following demographic descriptors were the independent variables: (1) Age of institution; (2) Ratio of Graduate Student Headcount to Total Student Headcount; (3) Ratio of International Graduate Student Headcount to Total Graduate Student Headcount; (4) Ratio of Graduate Degrees Awarded to Total Degrees Awarded; (5) Ratio of Science & Engineering Graduate Degrees Awarded to Total Graduate Degrees Awarded; (6) Percent Full-Time Faculty; (7) Percent Part-Time Faculty; (8) Ratio of Full Time Equivalent Tenure/Tenure-Track Faculty to Total Full Time Equivalent Faculty; (9) Ratio of Part-Time Faculty Headcount to Full-Time Faculty Headcount; (10) Full Time Equivalent Student to Full Time Equivalent Faculty Ratio; (11) Ratio of Total Research Space to Total Space; (12) Ratio of Total Library Volumes to Total Student Headcount; (13) Ratio of Total Graduate Semester Credit Hour to Total Semester Credit Hour; (14) Total Number of Nobel Prize Winners and Members of National Academies; (15) Hispanic Serving Institutions; (16) South Texas Border Initiative Schools; (17) Federal Research Expenditures; (18) Total Research Expenditures; (19) Tenure/Tenure- Track Full Time Faculty Equivalent with Teaching Responsibility. The only significant institutional demographic variable was the ratio of total parttime faculty headcount to total full-time faculty headcount. Those institutions with more part-time faculty may be able to alleviate teaching or other duties assumed by full-time faculty, allowing them to concentrate more on building research capacity. In addition, as the rate of growth of non-RDF funds between the first fiscal year of RDF (2002) and the latest year available (2007) increases, so does the ratio of part-time faculty headcount to total full-time faculty headcount. Eventually, there could be diminishing returns. If an institution’s part-time faculty grows at too great a rate compared to its full-time faculty, a decrease in research capacity could result.

Additional Analyses

Additional analyses were done to determine if relationships existed between the rate of change of research capacity and a related set of 2007 institutional variables. These analyses attempted to determine if a change in research capacity might have impacted other seemingly related institutional outcomes. Pearson r correlations were computed using rate of change as the independent variable and selected institutional demographic variables as the dependent variables. These analyses further questioned whether there was a relationship between growth in research capacity and some variables of interest. In other words, did research capacity influence current situations in fiscal year 2007? The results indicated that rate of change in research capacity did not influence current situations in fiscal year 2007. However, there could be other variables that were not identified and analyzed that could have influenced the current situations in fiscal year 2007. For fiscal years 2002, 2005, 2006, and 2007, the THECB required all universities participating in the RDF to submit annual spending reports on their RDF appropriations. The annual reports included RDF expenditure amounts and details by project or initiative, and revealed the following items of interest: (1) the number of projects reported by each institution increased each year; (2) the total RDF appropriations reported by all participating universities each year did not match the total RDF appropriations distributed for each of the reporting fiscal years (universities have the option to roll over or carry forward RDF appropriations between fiscal years within bienniums); (3) universities have used the majority of RDF for research projects; (4) in fiscal year 2002, the first year of the RDF, the second most common RDF expenditure was on laboratory-startup, followed by research administration; for the remaining fiscal years, 2005, 2006, and 2007, the second most common RDF expenditure was on student assistantships and other student-related expenses (the majority of expenditures for students during these fiscal years were reported by one institution, the University of Houston); (5) the majority of RDF appropriations were used for science, mathematics, and engineering disciplines in all fiscal years reported. Further analysis was conducted between the research/doctoral institutions and the masters/bachelors institutions. Between fiscal years 2003 and 2006, the research/doctoral institutions had a 17.19 percent increase in restricted research expenditures, compared to a 47.28 percent increase at masters/bachelors institutions. Comparing this with the 21.30 percent change in total federal research expenditures among all U.S. colleges and universities for the same years (NSF, 2007), it appears that the smaller schools are finding the Texas RDF to be making a positive impact on their research capacity.

Conclusion

This study showed a positive impact of the Texas Research Development Fund among its participating Texas public institutions, both for the research/doctoral-intensive universities and for the masters/bachelors universities. If the trend continues, participating Texas universities may be on their way to a higher level of research capacity. Paul Romer, one of the primary developers of the New Growth Theory, held that ideas are goods that are produced and distributed in a similar way as other goods. He linked the neoclassical theory of inputs with technology (Romer, 1996). According to this theory, economic growth is sustained by the way societies deal with advances in technology. Romer contended that research and development activities, which are associated with innovation, are the driving force of long-term economic growth. With innovation, ideas are generated for new products or new processes. With new products and new processes, better quality goods are produced, raising productivity (European Commission, 2001). Using this logic, it may be concluded that increased research and research capacity in Texas can result in economic growth for the state.

With direct access to patients and clinical investigatorswho fuel and test newideas, academicmedical centers (AMCs) play an important role in innovation that improves patient care and health outcomes [1]. However, an increasingly complex environment for clinical research in healthcare delivery settings could threaten this path to discovery if inefficiencies are not addressed [2–4]. Challenges facing clinical investigators include lack of time for research due to demands of clinical practice; disruption in established workflows associated with a transition to electronic medical records; insufficient infrastructure; and expanding complexities of data collection, government and institutional regulation, and burdensome contract negotiation [3,5–7]. As summarized in a 2015 National Academies report, “continuing expansion of the federal regulatory system and its ever-growing requirements are diminishing the effectiveness of the nation's research investment by directing investigators' time away from research and training toward overlapping and incongruent administrative matters.” [8] Many associate an increase in administrative burden with discouraging clinicians from pursuing a career in research, resulting in a shortage that could jeopardize the clinical research enterprise [4,7,8]. These challenges underscore the need for expert assistance from clinical research administrative support so clinical investigators can focus on science that will lead to improved patient care and health outcomes. Compared with a traditional model in which the management of studies is performed in a coordinator/investigator unit responsible for all trial activities, there is a growing interest inwhether consolidating research administration and support activities within a department or institution may contribute to more effective infrastructure for clinical research [9].However, barriers to implementing unified support systems may include lack of faculty support due to perceived loss of autonomy in research conduct [10].Detailed accounts are needed of transitions to unified research support offices, along with data to evaluate the potential benefits and limitations of such a model. This article reviews the Duke University School of Medicine experience with developing a centralized, support-oriented approach to managing clinical research.

2. Case study: Duke Office of Clinical Research

2.1. Transition to a unified support office

The transition to a unified research support office occurred gradually. Because of the proliferation of one-off approaches to institutional support for study teams, Duke created infrastructure and policy changes to provide a more consistent oversight approach for site-based clinical research. The changes were originally carried out under the Vice Dean for Research, who oversaw both clinical and basic research. Duke first created oversight bodies in the disease-based clinical departments and institutes. These Clinical Research Units (CRUs) were initially focused on addressing industry-sponsored trial oversight andmanagement. Regardless of size, each CRU created a charter, named a director, research practice manager, and financial practice manager. The CRUs were supported by a narrowly scoped Clinical Research Support Office (Fig. 1A); this central administration office pushed business processes and responsibility to the CRUs and served as a pass-through checkpoint. Although there was a marked benefit in oversight and compliance at the onset (e.g., improvements in clinical research billing, increased focus on training), with time this model became problematic. There was a lack of consistent processes across CRUs, given that each could establish policies and procedures unique to the history of the unit. These inconsistent policies (e.g., differences in clinical research training requirements, variations in feasibility review) created confusion for faculty performing research in multiple CRUs. After years of mandates, poor customer service, faculty isolation, and inconsistent study management at the CRU level, the institution established the role of Vice Dean for Clinical Research to improve processes and communication involving the CRUs and central administration. Organizational changes included new reporting structures for the Institutional Review Board, Clinical Research Support Office, regulatory affairs (IND/IDE management), and research integrity. The key cultural change involved rebranding the Clinical Research Support Office by expanding the concept and scope of an established Research Management Team from the Duke University School of Nursing [11]. The Research Management Team originally provided flexible, cost-efficient institutional support for research coordination, data management, and informatics. Its “study teams as my customer” approach served as a proof-of-concept model for what would become the new Duke Office of Clinical Research (DOCR; Fig. 1B). In 2012, the ResearchManagement Team's leader was installed as the director of DOCR to reinvigorate research support under the new office. Key changes included bringing senior information technology expertise to bolster support for clinical research software applications and reporting [11], establishing a new position to liaise with contracts and finance departments, and creating designated staff to handle ClinicalTrials.gov registrations and reporting [12]. DOCR increased from 14 to 54 full-time staff members over 3 years and now provides a broad range of support services (Figs. 1B, 2). Some research support remains within the individual CRUs, which vary in size and scope, but the strengthened unified support office provides access to consistent navigation, tools, and training for the N700 clinical research coordinators and other research support staff throughout the institution. DOCR's mission is to improve patient care through outstanding clinical research that is supported by sustainable business processes and a well-trained clinical research workforce.

2.2. Services

DOCR services span the stages of a clinical study and are supported by staff conducting a range of activities, including study startup, research builds in the electronic health record, and outreach andmentorship (Figs. 1B, 2). Studies that will be billing any research costs through the Duke University Health System are required to have a study initiation meeting with DOCR to review the protocol for clinical and research-specific activities, identify needs for the electronic health record (e.g., order sets, billing calendar), review qualifying status and national coverage decisions, and identify funding for activities and coverage of related activities in alignment with health system financial policies. Use of all other DOCR services is voluntary. During study planning, investigatorsmay receive a free consultation (subsidized by the School of Medicine and Duke Clinical and Translational Science Award [CTSA]) to determine project needs, and an estimate for DOCR services is provided. In this way, DOCR is able to tailor support to the needs of the individual study and research team. For example, experienced and well-staffed study teams conducting industrysponsored studies may need little assistance from DOCR. In contrast, an investigator-initiated NIH-funded study is likely to have a different set of needs. Aside from the initial study planning meeting, requests also may be submitted at any time through a central DOCR email address. DOCR publicizes its services on its website, through regular presentations, and via interactions with study teams (i.e., word of mouth). As a unified research support office, DOCR is able to offer a wide range of services. Many DOCR staff are cross-trained, enabling an efficient infrastructure that can meet the objectives of multiple research programs. Some investigators have indicated that they would not have felt comfortable supporting full-time research administrative staffwithin a small group due to uncertainties of future funding. By providing partial effort with specific spheres of expertise, we hypothesize that investigators will be less likely to assign tasks to an employee with insufficient background (e.g., a coordinator assigned the additional role of data management). This DOCR shared service model allows researchers to more tightly manage labor expenditures by only paying for services when they are needed. The shared-effort coordinator and data management pools reduce pressure frominvestigators and provide job security for highly skilled staff,which allows the institution to develop and retain top talent.Workloads can be adjusted to prevent burnout, a common problem encountered with research coordinators in AMCs [13]. Additionally, managing staff can relieve faculty from the burden of addressing any performance issues. Finally, this service can provide a quick stopgap in the event that a research team unexpectedly loses some of its staff during a study. This staffing model therefore benefits both the researchers and administration. In general, requests for datamanagement or clinical research coordinator effort are submitted via a Research Electronic Data Capture (REDCap) survey for that purpose. Requests also may come in via direct contact with a DOCR director, associate director, or Research Management Team member. A DOCR associate director contacts the study team to assess the effort needed, records the information in the DOCR effort database, and generates an agreement to send to the study team. DOCR provides free guidance on study design, protocol writing, and institutional reviewboard submission to all Duke residents and fellows. We believe these services help establish positive relationships with DOCR early and will ultimately lead to higher quality research conduct at the institution. The theme of facilitation and navigation is central to DOCR services (Fig. 2). It liaises with the CRUs and other institutional offices to shepherd studies throughout their life cycle. Close ties at the leadership level with the School of Medicine Finance Office, Office of Research Informatics, and Institutional Review Board, among others, help maintain productive partnerships (Fig. 1). Parallel processes are initiated whenever possible to avoid delays; for example, an operational review can be conducted concurrent with institutional review board review and contract negotiation. Even where support services reside within other groups, DOCR strives to add value through coordination and communication. A communications teamproduces newsletters on relevant topics for faculty, research staff, and CRU leadership; topics are gathered from all research administrative offices. Furthermore, DOCR attends regular meetingswith stakeholder groups to engage around initiatives, policies, or procedures that may involve multiple stakeholders. 2.3. Operations

DOCR shares an administrative manager with the Office of Research Informatics, who assists with budgets and finance. Projects and effort are tracked in an internal database, and staffing is increased as needed to keep up with demand and prevent burnout. When DOCR rebooted, CRU staff were not absorbed, but they were eligible to apply for open DOCR positions. Turnover has been minimal within DOCR, though some staff do leave or go on to other positions within the institution. DOCR is a leader in developing staff and encourages staff to consider well-matched opportunities across Duke.

2.4. Training/professional development

An important facet of the unified support office is to contribute to the growth of a network of research professionals. In 2015, DOCR offered N300 classes on topics including human subjects research, study documentation, and data security. There were over 13,000 individual completions across the in-person, blended, and online courses. DOCR not only offers training, but also an opportunity for advancement that was not previously available tomany clinical research staff. Skilled coordinatorsworking under a single principal investigator may not envision a path for professional growth within the institution, and may feel pressured by the investigator to remain stagnant within a uniquely defined position. By harmonizing research support, a new avenue for career development is opened, and the institution can provide support and structure to assist future generations of research professionals in becoming leaders who are truly invested in team research. Potential for career progression also helps establish clinical research professionals as leading practitioners in the clinical research industry. In support of this, DOCR currently is leading a project to modernize all job descriptions for research staff in the School ofMedicine. Job levels will be standardized across the institution and based on education, experience, and workforce competencies, and will allow for individual professional growth [14]. In concert with this effort, human resources is conducting a market analysis of salaries. DOCR also launched a voluntary Research Professionals Network at Duke in 2014 that so far has N400 members. The network connects research professionals fromacross the institution to advocate for research careers, provide options for formal and informal education, and develop high standards in the research community.

3. Results

3.1. Efficiency

Although no specific efficiency goals were established when DOCR rebooted clinical research support, the office has been tracking data and monitoring improvements to establish ongoing metrics. When focusing on studies that were required to have a study initiation meeting with DOCR, decreases in the institutional review time are not yet demonstrated, but a decrease in the time from institutional approval until the first participant is consented has been observed (Table 1). By continuing to provide expert navigation and better prepare research teams, DOCR aims to continue to improve these timelines. Importantly, there has been a reduction in the percentage of these studies that close without consenting any patients (Table 2), which could signify a more effective reviewprocess that is selective of research likely to be successfully implemented. The overall number of clinical research studies, average number of patients accrued per enrolling study, and total number of patients enrolled across the institution increased over time; the types of studies remained constant (Table 3). During the period from 2011 to 2015, the percentage of patients seen in the health system who were enrolled in clinical research increased from 1% to 7%; this exceeded the goal of increasing enrollment by 5%. One of the ways DOCR has increased efficiency is in the contracts process. A common complaint heard by School of Medicine leadership fromfacultywas howlong it took for a contract to be negotiated and executed at Duke. DOCR hired a contracts liaison to research this issue and help navigate contracts through the process. After meeting with representatives fromall CRUs, a common complaint emerged—the time from completion of negotiations to getting Duke's institutional signaturewas too long and not transparent. To better understand this issue, the contracts liaison reviewed the average turnaround times to get institutional signature, which was obtained in a separate central office. By moving the signature function to within DOCR, the average time to get institutional signaturewas reduced from3.6 days to 0.12 days.While the original 3.6 days may not seem like a large delay, the improvement shows that therewere inefficiencies in the process. DOCR streamlined required components and made it a priority to get the contracts signed within 24 h of receipt. Additionally, DOCR provided frequent status updates if there was a contract that could not be signed within 24 h and worked with the team to shepherd it through the process and get it approved for signature. This focus on turnaround time and customer service has eliminated the complaints regarding the long wait for institutional signature. The contracts liaison continues to identify other areas of the process that delay contract negotiation and execution and is working through each one to minimize inefficiencies and make the process as transparent and user-friendly for study teams as possible. Compliance with results reporting in ClinicalTrials.gov is another area where DOCR has shown improvements. On-time reporting has increased from b50% to 90% to 100% compliance over the past year.

3.2. Satisfaction

Satisfactionwith DOCR services is assessed by an annual survey sent to individuals on the DOCR mailing list. The satisfaction rates have remained at approximately 75% or higher throughout DOCR's history (Fig. 3A). Similarly, the proportion of negative comments received stayed relatively constant over time, with slightly more negative comments in fiscal year 2014, perhaps corresponding to frustrations associated with the implementation of the electronic health record across the health system (Fig. 3B). In addition, satisfaction for Research Management Team services (data management and clinical research coordinator support) is assessed via email or phone call after one month, and then using a feedback survey sent after the effort has ended. Ratings for this support over a 1-year period have been very positive (Fig. 3C, D). In fiscal year 2015, of individuals evaluating their optional study planning consultation (N = 16) and grant development assistance (N=12), 94% and 100%, respectively, indicated theywould be extremely or quite likely to use the service again. DOCR aims to hire high-performing staff and match appropriately skilled and experienced individuals to study needs. If a performance issue or mismatch occurs, this can be corrected quickly for the study team with a replacement from DOCR, generally within 1 week. In this model, investigators and study teams do not have to spend time hiring or addressing performance issues themselves.

3.3. Cost

The operating budget of the narrowly scoped Clinical Research Support Office prior to transition to DOCR was approximately $916,000 in fiscal year 2012. In fiscal year 2015, DOCR had an operating budget of $2.68 M, which included $905,000 covered by the CTSA grant. DOCR is paid for as part of the indirect cost rate for clinical research, which has not changed as a result of DOCR's inception. Only direct effort costs are charged to studies; there is no charge for overhead,management, facilitation, or consultation. In 2015, DOCR's Research Management Team had gross costs of $1.3 M, all of which were covered by research grants or contracts. From fiscal year 2012 to 2015, the institution's costs for clinical research compliance and auditing (external to DOCR) remained virtually unchanged, from $2.48M to $2.45M, not accounting for inflation. This was in spite of a growing number of research studies and patients enrolled (Table 3).

4. Discussion

Evenwhen the need for research administration is recognized, there can be disagreements over the structure of support and responsibility for oversight. We have observed that the concept of “central” support can be met with a negative reaction—the perception being that these systems are heavy-handed and inflexible, consisting of top-down policy and bureaucracy. Conversely, in our experience supporting research teams, individualized approaches to research conduct throughout an institutionmay appear effective, or at leastmore expedient, but frequently lead to operational, regulatory, and compliance challenges. In a transition to the unified support office, we observed that a culture change from within research administration may help reduce barriers to acceptance. To avoid creating infrastructure for the sake of infrastructure, it was essential for DOCR to apply its efforts with the mindset of optimizing clinical research while also protecting human subjects from potential harmdue to poor research conduct. Assessment and reassessment of existing administrative processes was essential to this process. Without thoughtful, multi-source input, development of institutional policies and procedures may distract faculty from research conduct and create a perception that administration is hampering research progress. A focus throughout DOCR's growth and development was on building relationships with clinical investigators and study teams. Demonstrating advocacy for researchers by finding solutions that are compliant with policies and regulations while allowing the research to proceed expediently is the currency to support cultural change. Support from senior leadership within the institution was also critical throughout the office's transition. By assuming the role of facilitator, DOCR empowers staff within study teams to speak up when an issue needs to be addressed. This may not always be the case in decentralized structures where a principal investigator is also the supervisor of an administrator or coordinator. Other consolidated units supporting clinical research at AMCs have been detailed in the literature, but those have been typically implemented at the level of a department or group of departments [15,16]. This paper describes experiences implementing a clinical research support office for an entire AMC. As with DOCR, efficiencywas described as a major goal of the previously published examples, though data to support the efficiency of such a model remain limited. In an analysis of high-performing clinical research teams at research centers by Retsch- Bogart and colleagues, characteristics included shared efficient processes, continuous process improvements, and a business-like approach to clinical research [17]. All of these are approaches employed by DOCR. Moving to a unified research support office involved an increase in cost to the institution, which was expected, and has been described with transitions at other institutions in association with increased services [16]. However, during a period in which studies and enrollment increased, compliance costs remained stagnant. If DOCR can prevent compliance issues through training/education, consultation, and its suite of support services ensuring high-quality research, this will be of value to the institution. While some institutions approach shared support as a cost-savings measure, another view is to recognize research administrative support as an investment,with returns in the formof increased collaboration and retention of clinical investigators through the delivery of high-quality services [18]. A limitation in our analysis is that formal satisfaction data are only available starting after the establishment of DOCR. Improvements in study startup times have only begun to be observed, andwewill continue to monitor these and other study metrics to gain a better understanding of the effects of the support office. Rollout of a clinical research management system within the next year will provide access to more data for assessing changes in efficiency and return on investment for DOCR. It is unknown how long it would take to observe such changes.DOCR, alongwith CTSA administrators, isworking on assessing common metrics, both within and across institutions [19]. Currently, this remains challenging with varying systems to collect and analyze data. As new electronic systems (such as the clinical research management system) are deployed, these common data elements can be captured to standardize data collection across the institution for all studies.

4.1. Remaining challenges

While significant strides have been made to improve clinical research oversight and operational support at Duke, there is much to do. In some areas, trust between faculty and administration must be reestablished due to previously over-burdensome processes. Some investigators prefer to maintain their individualized approaches to conducting research. Having many of DOCR's services categorized as optional allows some flexibilitywhile still providing overarching structure and support. Since the office evolved gradually, there is still some duplication with remaining CRU support, but this is expected to shift over time. Keeping focused on the ultimate goal of improving patient care and continuing to foster collaborations will help administration and faculty move forward as a team. Shifting a decades-long investigator/coordinator research culture to this collaborative approach between central administration and a study team requires communication, trust, and a willingness to accept component responsibilities at all levels of a research organization. As the role of research professionals is increasingly professionalized, it should bemorewidely recognized that helping navigate the research environment is a valuable, high-level skill. In addition, DOCR,while successful in providing an increased range of services,must be mindful to grow responsibly and not just for the sake of growth. Finally, wemust continue to share experiences to cultivate more effective research administration, particularly throughout the CTSA institutions, to transform the clinical trial enterprise.

5. Conclusion

The Duke Office of Clinical Research provides one example of how unified research support can be implemented within a large AMC. Although unified research administration and support is often met with negative reactions initially, we have shown that it is possible to develop a team that cares and delivers excellent customer service to overcome this perception. The unified model has multiple advantages that can be realized through a shift in culture to refocus effort on the efficient and safe conduct of research that can lead to improved patient care.

Conflict of interest

In the last three years, Mark Stacy, MD, received grant funding from the Michael J. Fox Foundation, NIH, Parkinson Study Group, and Pharma2B. He has served in a consultation capacity for Acorda, Genzyme, Eli Lilly, Lundbeck, Pfizer, ProStrakan, SK Life Science, and Vanda. He has also served on data management or protocol steering committees for Allergan, Biotie, Merz, Osmotica, and Revance. Lindsey L. Spangler, JD, owns stock in Allergan. None of the remaining authors has any potential conflicts of interest to disclose.

Government data refers to the collection of information and datasets that are generated and maintained by governmental organizations. It includes various types of data such as demographic statistics, economic indicators, public health records, environmental measurements, etc. Open government data (OGD) refers to the government data that are published and made accessible to the public, in a format that is easily discoverable, downloadable and readable, allowing users to freely manipulate, analyze, reuse and distribute it for various purposes (Kassen, 2020). One of the primary goals of OGD is to promote public engagement and innovation. By making government data open, it enables citizens, businesses, researchers and developers to analyze and utilize the data to conduct research and create innovative ideas, as well as improve public services and hold governments accountable (Zuiderwijk et al., 2014). However, substantial studies on OGD indicate that the usage of OGD is lagged; even though a huge number of datasets are available, only a small number are actively used (Zuiderwijk and de Reuver, 2021). And the OGD-driven innovation only exists sporadically in private organizations and government-led hackathons (Mu and Wang, 2022; Susha et al., 2015). Scholars attribute this low level of data usage to underdeveloped institutions for data publication (Kassen, 2018; Li and Chen, 2021). The key proponents of institutional theory argue that the adoption of OGD is influenced by existing institutional arrangements and that proper institutional design will contribute positively to the transition from closeness to openness (Altayar, 2018; Safarov, 2019). For instance, based on the system theory, Janssen et al. (2012) explain that the conventional system boundaries of governments are dissolving and becoming open when data is made public, which requires different steering institutions to manage data, motivate data usage and stimulate system feedback from external stakeholders. Besides, from the human-data interaction perspective, Victorelli et al. (2020) perform a comprehensive evaluation of the literature on human-data interaction and conclude that institutions on data representation, data interaction and data processing should be formulated in advance to facilitate users to understand and interact with the government data and to take the best use of the data. Although scholars have reached the consensus that institutions are needed for publicizing data, they do not elaborate on what institutional dimensions should be considered and how these institutional dimensions will affect innovation? Do these dimensions work individually or together? Without answering these questions, people merely know institutional design is important but will never know what institutions should be at play and how they take effect. As a response, this article distills three institutional dimensions from the existing literature: the instructional, structural and accessible dimensions. And we examine how these institutional dimensions will affect Scientific Research Innovation (SRI). SRI refers to the process of formulating novel research questions, creating and developing new methods and tools and exploring new knowledge and theories. We choose SRI as the outcome variable because researchers, among the various stakeholders, are one of the largest groups whose innovation activities may highly depend on government data and thus be affected by data presentation (Lassinantti et al., 2019; Lnenicka et al., 2022). For example, Li et al. (2019) demonstrates that tabular and graphical data presentations may affect how researchers understand an under-researched area and subsequently influence how they formulate research questions. Again, Harron et al. (2017) argue that effective data linkage environments can facilitate researchers to answer questions that require large sample sizes or detailed data on hard-to-reach populations and generate findings with a high level of external validity and applicability for policy making. Therefore, our main research question reads as:

RQ1. How do the instructional, structural and accessible institutional dimensions, affect scientific research innovation in singular and binary forms?

Additionally, the literature suggests that data adoption and OGD-driven innovation may hinge on users’ characteristics. For instance, Venkatesh and Davis (1996) point out that only when an individual believes that the data is relevant and will assist him or her in improving job performance will he or she adopt it. In addition, Janssen et al. (2012) put forward a myth of OGD and explain that only the individual who has the resources, expertise and capacities to collect and process data, will use the data. In line with these thoughts, we introduce two mediating variables, researchers’ perceived data usefulness and data capability, the former measuring researchers’ subjective perception of data usefulness and the latter measuring researchers’ objective abilities of data collection, process and interpretation. Thus, this article addresses an additional research question:

RQ2. Do researchers’ perceived data usefulness and data capability play mediating effects between OGD institutions and SRI?

The empirical evidence of this study comes from the Chinese context. China is a typical case of a transition economy that performs well in scientific research innovation but institutional construction for OGD has just started. Therefore, to sustain and even boost scientific research innovation and to inject new data-based dynamics to innovation, the establishment of proper institutions for data publication are imperative. For other transition countries as well, a broad range of institutional changes and design work are needed to release data value and specially to promote effective data-driven innovation. However, the OGD institutions in western, developed countries might not fit the socioeconomic contexts of the transition economies. Thus, China, as a representative case, need to consider what the institutional dimensions, in singular and binary forms, can set the stage for OGD usage and engender positive societal impact, such as scientific research innovation in our case. This study has both theoretical contributions and practical implications. Theoretically, the study compensates for the theoretical gap between OGD institutions and SRI. In addition, it examines the role of researchers’ perceived data usefulness and data capability in linking OGD institutions and SRI. In practice, the conclusions on the influence of OGD institutions on SRI provide a timely reminder for policymakers, administrators and practitioners to manage OGD.

2. Theoretical background

2.1 Institutional dimensions of open government data

Although institutions are defined in many ways, they are generally understood as formal rules (e.g. laws, regulations, policies, standards, or guidelines) and informal norms (e.g. cultures, customs, or traditions) that constrain and encourage individual behaviors and social, political and economic interactions (North, 1991). In this study, we only consider formal institutions because we aim to analyze how formal rules for data publication influence innovation; informal institutions are not considered because we do not attempt to explore the influence of cultures and norms on data usage. Three fundamental institutional dimensions that would affect data usage and influence scientific research innovation were extracted from the extant OGD literature (Altayar, 2018; Janssen et al., 2012; Li and Chen, 2021; Machova et al., 2018; Victorelli et al., 2020): (1) the instructional dimension; (2) the structural dimension; and (3) the accessible dimension.

2.1.1 The instructional dimension: whether governments should provide instructional and security rules for data usage, or just publicizing data without any instructions or monitoring?. One of the dominant debates about OGD is whether governments should publish data usage instructions and security rules along with the opened data. The current literature on data’s instructional dimension does not offer an explicit answer. Institutionalism theorists assert that the failures of the human-data interaction are the consequence of inadequate rule design that limits the users’ understanding of the opened data and the data context (Gonzalez-Zapata and Heeks, 2015; Niebel, 2021). Take this regard, Wang and Lo (2016) argue that data providers should be responsible for giving data instructions (e.g. data provenance, suggestions and requirements on how to use the data and recommended processing software) along with data publication. As a supplement, Bonina and Eaton (2020) recommend that, except for instructional rules, the arm’s-length connection between data providers and data users should be governed by security rules to protect data use.

On the contrary, some scholars argue that too many data instructions and security rules will constrain data usage and exploitation. For example, Martin et al. (2019) points out that in some circumstances data instructions and security rules may hinder innovation because the innovators may (1) abandon the innovation ideas to focus on others that face fewer data regulations; (2) feel instruction burdensome and be discouraged from using the data and (3) minimize data usage and thus reduce their attempts to access data. Similarly, Niebel (2021) reports that data protection rules have a negative impact on innovation because innovators must adhere to the security standards, which will increase the costs of innovation. 2.1.2 The structural dimension: whether governments should provide highly structured data or just the data in loosely structured formats?. The second debate centers around the structure of data, that is, whether governments should supply highly structured data or just drop it and do not bother about the format. A specific response is not provided by the literature currently available on data’s structural dimension. According to Kitchin (2014), structured data “are those that can be easily organized, stored and transferred in a defined data model, such as numbers/text set out in a table or relational database that have a consistent format.” In contrast, unstructured data do not have a predefined data model or common identifiable structure, such as narrative text, audio, photo, or video. Some scholars argue that governments are better to publish structured datasets because such data are “machineprocessable”, meaning that calculus and algorithms can read, combine, process and analyze them easily, and computers can depict them using graphs and maps (Attard et al., 2015; Zuiderwijk and Janssen, 2014). In the field of scientific research, Figlio et al. (2017) propose that the integration of structured datasets not only provides researchers with a full-sample data resource that reduces the generation of random errors during empirical analysis, but also offers new opportunities to reveal the full picture of event development under dynamic longitudinal panel data. However, other scholars point out the problems of highly structured data. Grossman and Pedahzur (2020) argue that in the big data era, only 15–20% of existing data are structured data, while most available data are unstructured, including political speeches, pictures, video recordings, media broadcastings, policy/regulatory documents and massive blog posts generated by the wider public. And these unstructured data are growing much faster than structured data (Zikopoulos et al., 2012). For these data, governments are not advised to open them in a highly structured format, since data structuring process is a reductive process that inevitably entails the loss of details and context, and the structuring process may not keep the pace of data generation (Grossman and Pedahzur, 2020). These data are qualitative in nature and human-readable. Therefore, users can convert the unstructured data into structured data, depending on the users’ own needs and purposes, through imposing a common structure upon the data by classification and codification. 2.1.3 The accessible dimension: whether governments should set up data access requirements or let public access data directly and effortlessly?. The third debate focuses on the data accessibility issue, that is, whether governments should set up certain data access restrictions (e.g. registration, application, or payment) or anyone is able to obtain the data without any additional efforts. There is no clear solution provided by the literature currently available on data’s accessible dimension. Some scholars argue that complete and immediate disclosure of government data is needed, and the insurance of free public access is considered a significant foundation for open government and transparency (Dawes, 2010; Lourenco, 2015). For instance, The Open Government Working Group (2007) emphasizes several important access principles for government data, including complete (all government data should be made available online), accessible (data can be obtained directly, not through navigating web forms or additional technical tools), non-discriminatory (government data do not require registration and application) and license-free (government data does not have any copyright and thus data access should be free of charge).

However, other scholars challenge the legitimacy of the above-mentioned claims. For example, Janssen et al. (2012) argue that a greater amount of opened data does not necessarily lead to better data usage and exploitation. Data publication without prior screening can result in information overload at the societal level and lead the public to less understanding and more confusion of government data. In line with Janssen, Wang and Shepherd (2020) evaluate the UK’s OGD practice and point out that most published data hardly attracted public attention and the British government even halted the commenting and forum functions due to inactive citizens. In addition, Janssen and van den Hoven (2015) state that governments always need to consider the privacy and security issues when releasing data and thus will inevitably require users to register for data access and usage. In some circumstances, governments use the “disclosure upon application” strategy, meaning that, in order to avoid information overload and reduce privacy risks, users need to apply for permits/licenses to access certain needed data (Lnenicka and Komarkova, 2019). In summary, we can see that the current literature on OGD has recognized the institutional aspects of data publication. However, disagreements still exist on how to design the rules, i.e. whether instructional rules should be presented along with data publication; whether rules should stipulate data structure; and whether accessibility should be restricted to a certain degree. Our study aims to fill the gap by investigating the effects of both sides of each institutional dimension on scientific research innovation and thus unpacking the logic of how the institutional dimensions of OGD influence data usage.

2.2 Scientific research innovation

Currently, there is not a unified term to describe the “newness” produced from scientific research activities. The most influential term is “scientific creativity” (Simonton, 2003), which means the capacity of researchers to conduct scientific studies that are novel, original, valuable and unexpected. This definition focuses on creative individuals and emphasizes the individuals’ mental processes and cognitive operations that lead to scientific discoveries, but does not pay much attention to creative products. As a response, some scholars depart from the Schumpeterian tradition and use the term “research novelty” to describe creative products generated from scientific research (Lee et al., 2015; Schumpeter, 1934). Here, creative products do not only generate from scratch but also from the unprecedented and distant combination of existing bits of knowledge (Wang et al., 2017). Thus, this definition emphasizes recombination of existing knowledge and does not capture the creative products that are made from new and fresh inputs of research data. In this study, we borrow in the innovation theory (the data-driven innovation literature in particular) which argues that innovative products can be designed not only through existing knowledge but also through new analytical and productional materials such as open data (Jetzek et al., 2014; Rizk et al., 2022). Therefore, we build upon the concepts of “scientific creativity” and “research novelty” and introduce the term “scientific research innovation” (SRI) as our outcome variable of OGD’s institutional construction. It generally refers to the scientific research products that are “new” or contain “newness” in relative to the existing knowledge. The new research products can be novel research questions, inventions of research methods, instruments and tools and discoveries of new relationships between variables to reveal what we otherwise had not known or conceived (Corley and Gioia, 2011; Heinze et al., 2007).

2.3 Perceived data usefulness and data capability

In this study, we propose and examine two mediating variables linking the institutional dimensions of OGD and scientific research innovation. One mediating variable is researchers’ perceived data usefulness, which reflects the researchers’ subjective perception of OGD. It measures the extent to which the researchers think that OGD would improve their opportunities to generate SRI (Yoon and Kim, 2017). If the researchers believe OGD is useful and valuable, then they will put efforts into consciously collecting and analyzing the data, which lays the foundation for SRI (Weerakkody et al., 2017). By contrast, the researchers are likely to avoid finding and using OGD if they think OGD is useless. As such, we propose that perceived data usefulness plays a mediating role between the institutional dimensions of OGD and SRI. Another mediating variable is researchers’ data capability, which reflects the researchers’ objective aspect of OGD. It refers to the abilities or skills of the researchers to collect, translate, convert, analyze and exhibit data (Li et al., 2019). When the researchers have sufficient data capability, they know what they need, why they need the data and how to create the interfaces and systems of data (Rizk et al., 2022). That means, data capability can help the researchers understand and digest a large amount of data. Meanwhile, strong data capability facilitates researchers to explore and maximize the growth of data value by using proficiently appropriate methods and uncovering valuable information and patterns. Hence, if the researchers do not have the capability to use OGD, the new ideas, insights and designs with the data might be forgone and SRI becomes bleak (Jetzek et al., 2014).

3. Methodology and data

3.1 Questionnaire development and variable operationalization

This study relied on the following procedures to develop the survey questionnaire. First, we created a set of initial items from the literature. Second, we invited 15 experts from different disciplines to check the wording of the item and received 11 suggestions for revision. The comments from these experts are listed in Table A1. Based on these suggestions, we redefined the scope of OGD, revised the description of item IRA3 and added self-report items in the questionnaire. Next, we employed the pilot study that re-distributed the questionnaires to 144 respondents to further examine the quality of the questionnaire. Then, we evaluated the discriminant validity and the convergent validity of the questionnaire, removed the problematic items and finally obtained the formal 24-item questionnaire. The detailed items and their sources for the questionnaire are presented in Table A2. In the questionnaire, we added researchers’ gender, age and dependency of discipline on OGD as control variables. Furthermore, we regarded the level of research institute with which researchers are affiliated as an additional control variable (Lee et al., 2015), because in China research institutes are ranked, for instance, into the “985” and “211” series. In our models, the first-level research institutes include the Chinese Academy of Science (CAS) and “Project 985” universities [1], the second level includes “Project 211” universities [2] and the third level includes all other universities. We measured the dependent and independent variables using a five-point Likert scale ranging from 1 (i.e. strongly disagree) to 5 (i.e. strongly agree). The control variables were measured by category.

3.2 Sample and data collection

From August to September 2021, this study collected data through Wenjuanxing (www.wjx. cn), a popular and most prominent online survey platform in China. The respondents were researchers who used OGD in their research, including Ph.D. students, postdoctoral fellows and research fellows. All participants were recruited by posting the survey recruitment information on the crowdsourcing platform, such as muchong.com, which contains a large number of active researchers from multiple disciplines. Finally, a total of 1,611 questionnaires were received, of which 1,092 valid surveys were employed for this study. Table 1 shows the respondents’ demographic information. Among the 1,092 respondents, 56.96% are male and 43.04% are female. Most respondents are 21–30 years old, up to 82.23%. A majority of respondents come from the first-level research institutes (75.00%), while 7.97% are from the second level and 17.03% are from the third level. Regarding the dependency of discipline on OGD, 52.75% of the respondents agree that their research disciplines rely on OGD, while 13.09% disagreed and 34.16% of respondents are uncertain.

3.3 Strategies for data analysis

First, partial least squares structural equation modeling (PLS-SEM) was used to examine the direct and indirect effects of individual institutional dimensions on SRI. We chose PLS-SEM to test the relationship because the PLS-SEM method has several advantages over other statistical methods used for structural equation modeling (Hair et al., 2019). Most importantly, the PLS-SEM method is highly predictive and appropriate for research where the goal is theory development rather than theory testing. It is appropriate for our study because previous research has not extensively investigated the effects of the institutional aspects of OGD on innovation, and factor loadings and cross-loading external models allow the PLSSEM method to predict and explain such underdeveloped causal relations. In addition, research models with complex latent variables that are measured by multi-layer constructs can be predicted using the PLS-SEM approach. We find that the three institutional dimensions are measured by complicated structures of items, making this method an appropriate analytical approach for our study. Furthermore, our study involves a formative variable, the unrestricted accessibility rule, which requires the PLS-SEM method because this method permits formative variables. Second, polynomial regression with response surface analysis was applied to test the direct and indirect effects of binary institutional dimensions on SRI. Polynomial regression can provide more insights by evaluating the relationship between the interaction of two predictors on the dependent variable (Shanock et al., 2010). This allows for capturing more nuanced and intricate patterns in the data. It thus fits our research purpose of examining the binary effects of the institutional dimensions on SRI. Furthermore, polynomial regression is often used in response surface analysis, which offers us with visualized, easy-to-interpret non-linear relationships between the binary institutional predictors and SRI. Moreover, the coefficients of the polynomial terms provide insights into the direction and magnitude of the relationships between the binary predictors and SRI. This enables us to identify the optimal conditions or settings for the institutional dimensions that maximize SRI.

3.4 Testing the quality of the research model

To assess the internal consistency and reliability of the questionnaire, Cronbach’s alpha and composite reliability (CR) were used as indicators. Meanwhile, the convergent validity was assessed by the average variation extraction (AVE). Table 2 shows that all constructs’ Cronbach’s alpha coefficients are greater than 0.6, the CR of latent variables is greater than 0.7, and the AVE is greater than 0.5, indicating that all constructs in the research model are highly reliable and convergent (Chin et al., 2003). The discriminant validity of the square root of AVE was then tested using cross-loading analysis and the association of the square root of AVE with other components (Hair et al., 2019). Cross-loading revealed that all factor loadings are greater than the suggested value of 0.70. Table 3 further shows that the squared root of all variables’ AVE is greater than their correlation with other factors. As a result, there is good discriminant validity between variables. Then we calculated the variance inflation factor (VIF) to see if there was multicollinearity. If the value of VIF exceeds the 3.3 threshold, multicollinearity is a worry (Diamantopoulos and Winklhofer, 2001). The maximum value of VIF in our analysis is 2.44, indicating that multicollinearity is unlikely to be a problem. We also investigated the possibility of common method bias. First, SPSS 23 was used to run Harman’s single-factor test (Podsakoff et al., 2003). The findings revealed that the first component only explains 26.46% of the variation, falling short of the 50% requirement. Second, SmartPLS 3.3.3 was used to run the unmeasured common latent technique, which involves adding a major construct of all variables in the research model, as described by Liang et al. (2007). We found that the average factor loading value in the substantively principal constructs is substantially greater than the common approach when we compared the average variance of each item in the substantively principal constructs and the common method. As a result, the typical procedure bias would not be a threat.

4. Results and analyses

The results are presented in two parts. The first part shows how institutional dimensions affect SRI individually, as well as the roles of mediating variables in linking the institutional dimensions and SRI. The second part displays the binary effects of the institutional dimensions on SRI and the roles of mediating variables.

4.1 Testing the direct and indirect effects of individual institutional dimensions on SRI Table 4 presents the results of the direct effects of individual institutional dimensions on SRI. As can be seen, significant effects of instructional rules on SRI emerge (β 5 0.337, p5 0.000). That means, when governments provide instructions on data usage along with data publication, SRI is more likely to be produced. This result is consistent with the opinion of institutionalism theorists who argue that proper instructions on data usage will ease the interface between data and users and thus is advantageous to data exploitation (Mutambik et al., 2023). At the structural dimension, the results show that both unstructured data (β 5 0.078, p 5 0.023) and structured data (β 5 0.064, p 5 0.035) have significant effects on SRI. This finding is in line with what we argued in the theoretical framework: governments can publish both quantitative, machine-readable datasets in rows and columns and qualitative, human-readable texts, audios and videos. The reason is that structured and unstructured data have their respective advantages: structured data that is organized into a predefined consistent format is easier to store, search, retrieve and analyze and thus allows for efficient data processing and simplified data management; unstructured data can provide richer and diverse information, especially valuable contextual information and can offer researchers a more comprehensive and holistic view of the data and allowing for deeper analysis and understanding (Grossman and Pedahzur, 2020). However, at the accessible dimension, the result only supports the positive relationship between restricted accessibility and SRI (β 5 0.083, p 5 0.012); the effect of unrestricted accessibility on SRI is not supported. This indicates that the use of restricted data is more beneficial for SRI. This logic may be interpreted by the fact that restricted data ensures that sensitive or confidential information is already processed and protected by governments, reducing researchers’ risks of data misuse and unintended ethical consequences and increasing researchers’ trust on the data. Moreover, restricted data might be of higher quality because it often undergoes more rigorous quality control measures within governments, ensuring its accuracy and reliability for researchers’ usage (Meijer et al., 2014). In addition, among the control variables, the results show that research institute level and dependency of discipline on OGD are related to the dependent variable, while researchers’ gender and age are not. Apart from the direct effects, we also test the indirect effects of institutional dimensions on SRI through researchers’ perceived data usefulness and data capability. As Table 5 shows, there is a “complementary” mediating effect of perceived data usefulness between instructional rules and SRI (95% confidence interval (CI) 5 0.046–0.094). Meaning that instructional rules transfer their effects to SRI partly through perceived data usefulness. It suggests that data instructions can make researchers think the data is useful because they provide clear guidance on how to use and interpret the data effectively; when researchers perceive the data is useful or of high quality, they will be satisfied with the data and generate greater reuse intention (Wang et al., 2023). Besides, there is an “indirect-only” mediating effect of perceived data usefulness between restricted data accessibility and SRI (95% CI 5 0.009–0.045). It means that perceived data usefulness functions as a necessary condition for restricted accessibility to influence SRI. The underlying logic would be that when restrictions exist in data accessibility, researchers would perceive higher data security, have greater trust on the data and will attach more value to the data (Bargh et al., 2016). In that case, researchers will feel more confident to generate valuable new insights or knowledge. Regarding researchers’ data capability, it plays a partial intermediary role in linking instructional rules and SRI (95% CI 5 0.040–0.088). It suggests that clear data instructions and guidance (e.g. how to use the data and suggested analytical tools) will enhance researchers’ data capacity and further lead to innovative discoveries (Wilson and Cong, 2021; Li et al., 2019). However, it plays complete mediating effects between structured data and SRI (95% CI 50.009–0.036), unrestricted data and SRI (95% CI 5 0.003–0.029) and restricted data and SRI (95% CI50.010–0.042). These results indicate that the generation of SRI depends highly on researchers’ data capability when the data used are in structured formats. And no matter whether there are access restrictions or not, it requires data capability for researchers to produce SRI. Figure 1 summarizes the direct and indirect effects of individual institutional dimensions on SRI.

4.2 Testing the direct and indirect effects of binary institutional dimensions on SRI

In our study, we construct three models of binary institutional dimensions. Table 6 reports the polynomial regressions results as well as the slopes and curvatures along the congruence line and incongruence line for the three models respectively. Here, congruence means two combined institutional dimensions exhibit concurrently high-high or low-low statuses, while incongruence measures the opposite statuses, high-low or low-high, of the combined institutional dimensions. In the regression models, we controlled researchers’ gender, age, institute level and dependency of discipline on OGD. Additionally, when we combined two dimensions, we treated the other dimension as the control variable. Figures 2–4 illustrate the three-dimensional response surface based on the coefficients. First, in Model 1 (instructional and structural rules), Figure 2b shows a positive significance (slope50.380, p50.000) along the congruence line. It also indicates that the performance of SRI is higher at the rear corner (high instructional rules and high structured data) than at the front corner (low instructional rules and low structured data) (also see Figure 2a). The incongruence line also reports a positive significance (slope 5 0.378, p 5 0.000). Figure 2c shows that the performance of SRI is higher in the left corner (high instructional rules and low structured data) than in the right corner (low instructional rules and high structured data) (also see Figure 2a). When we compare the left corner along the incongruence line with the rear corner along the congruence line, we find that the rear corner brings about a slightly higher performance of SRI, meaning that the most powerful condition in triggering SRI in model 1 is the combination of instructional and structured rules. Second, in Model 2 (instructional and accessible rules), Figure 3b shows a positive significance (slope 5 0.468, p 5 0.000) along the congruence line. It also indicates that the performance of SRI is higher in the rear corner (high instructional rules and high accessibility) than in the front corner (low instructional rules and low accessibility) (also see Figure 3a). Along the incongruence line, a positive significance also appears (slope 5 0.282, p 5 0.000). Figure 3c indicates that the performance of SRI is higher at the left corner (high instructional rules and low accessibility) than at the right corner (low instructional rules and high accessibility) (also see Figure 3a). If we compare the left corner along the incongruence line with the rear corner along the congruence line, we find that the left corner brings about a higher level of SRI, meaning that the condition of high instructional rules and low accessibility is more beneficial for the generation of SRI. Third, in Model 3 (structural and accessible rules), Figure 4b shows a positive significance (slope50.089, p50.023) along the congruence line. It also shows that the performance of SRI is higher in the rear corner (high structured data and high accessibility) than in the front corner (low structured data and low accessibility) (also see Figure 3a). Along the incongruence line, however, the result reports a negative significance (slope 5 0.089, p 5 0.008). As Figure 4c shows, the performance of SRI is higher at the right corner (high accessibility and low structured data) than at the left corner (low accessibility and high structured data) (also see Figure 4a). However, this response surface does not show whether the rear corner condition or the right corner condition produces a higher level of SRI. Apart from the direct effects of the binary institutional dimensions, we also tested the indirect effects of the combined variables through the two mediating variables. As Table 7 shows, both perceived data usefulness and data capability play partial mediating roles between three pairs of binary institutional dimensions and SRI.

5. Conclusions and discussions

5.1 Summary of findings With a sample of 1,092 respondents in China, our empirical findings indicate that both the presence of data instructions and the presence of data access restrictions have positive impacts on SRI; however, it does not matter for SRI whether governments publish data in structured or unstructured formats. In addition, our findings also reveal that researchers’ perceived data usefulness plays a partial mediating role between the instructional rules and SRI; however, it plays as a necessary mediating condition between restricted accessibility and SRI. This implies that only when researchers perceive government data as extremely important, data accessibility restrictions would not limit data usage for SRI. Regarding researchers’ data capability, it plays a partial mediating role between the instructional rules and SRI; however, it plays a complete mediating role between structured data format and SRI. This implies that when data are presented in structured format, researchers need to depend on their data capability to interpret and process data and subsequently generate SRI. Furthermore, we investigated the effects of binary institutional dimensions on SRI to gain a better understanding of the intricate interplay between the rules and their interactive impact on SRI. As a result of Section 4.2, we can conclude that when governments provide instructional rules in data publication, restricted accessible rules or unstructured rules are preferred to go together to facilitate SRI. A possible explanation might be that, when governments put forward instructional and regulatory rules on data usage (e.g. for privacy and abuse-avoidance concerns), data acquisition is usually non-free and set with registration requirements. And to avoid losing data details and context information, governments usually publish unstructured data and in doing so, data usage instructions are preferably accompanied with the unstructured data to direct or guide data classification and codification. By contrast, unrestricted accessible rules must exist when governments do not provide instructional rules for data publication, and there are no requirements on data structure rules.This is possible because, when governments do not intend to impose any data usage instructions to further promote equal participation, any acquisition restriction rules such as registration and payment should be removed.And any data, no matter in structured or unstructured formats, should be encouraged to open to stimulate equal and easy participation and data usage.

5.2 Theoretical contributions

This work adds to the body of knowledge on OGD in the following ways. First, this study proposes an analytical framework of institutional dimensions of OGD, which argues that instructional rules, structural rules and accessible rules are three institutional pillars supporting OGD-driven innovation. This framework advances the current understanding of what kind of institutions will affect data usage and exploitation. As we explained in the introduction, although scholars recognize the importance of institutional design for OGD, they do not clarify what rules should be considered. Our framework just bridges this gap. Second, this study adds to the OGD literature by empirically testing the causal relationships between the institutional dimensions and scientific research innovation (SRI). The findings from the individual and binary effect of direct and indirect of OGD and SRI tests improve our understanding of the mechanism of OGD-driven innovation. Comparing our findings with the opinions in the OGD literature, the similarity is the discovery of importance of instructional rules and access restrictions to promote data usage and exploitation, but the difference lies in the fact that our findings do not support the significance of data format in influencing data-driven innovation. Additionally, and most importantly, we add the OGD literature by analyzing the interplay between different institutional dimensions and their joint effects on innovation. This is the new knowledge added to the OGD literature.

5.3 Practical implications

This research also has practical implications. These practical implications can be transferred to other transition countries that strive for advancing their scientific research innovation and building up suitable institutional frameworks for OGD. First, this study confirms the importance of instructional rules and accessible rules in promoting innovation. Therefore, it is essential for governments to develop and continuously construct instructions regarding data usage and to establish necessary data acquisition restrictions. In the field of scientific research in particular, governments also need to be aware that users’ perceived data usefulness and data capability will influence the performance of OGD-driven innovation. As such, governments are encouraged to carry out events that can enhance potential users’ perception on data usefulness (such as data marketing and educational activities) and that can increase users’ capability in using data (such as technical training activities). Secondly, our findings reveal the effects of binary institutional dimensions on innovation. A practical implication is that when governments provide instructional rules for data publication, restricted accessible rules or unstructured rules are preferable to go together to facilitate SRI. However, when governments fail to provide instructional guidelines for data usage, freely accessible rules must be available. And in this situation, there are no mandates for data structure rules, indicating that governments may or may not issue requirements on data structure.

5.4 Limitations and future research

There are three main limitations to this study that can be addressed in future research. First, this study relies on questionnaire data. This inevitably raises the subjectivity problems. In future research, objective data are encouraged to use to measure the variables. Second, our respondents are researchers in the Chinese research institutes, the scope condition for generalizing our findings is limited to the scientific research field. How the institutional dimensions should be designed to promote innovation in other fields, such as business models or public services, are subject to future research. Third, this study only considers researchers’ perceived data usefulness and data capability as two mediating variables, but more mediating variables, such as researchers’ social capital or organizational environment, should be tested in future research.

Promoting research integrity (RI) and responsible research practices is increasingly important to foster a healthy and responsible research climate (Science Europe Working Group on Research Integrity, 2016). When looking at the high percentage of researchers who admit research misconduct and questionable research practices (QRPs) (Fanelli, 2009; Gopalakrishna et al., 2022; Pupovac & Fanelli, 2015; Xie, Wang & Kong, 2021), one can understand why RI is gaining importance in science. RI intends to address the problem by performing research according to responsible research practices, and in line with high professional, methodological and ethical standards (Science Europe Working Group on Research Integrity, 2016). Raising RI awareness and promoting responsible research practices is a collective and multifactorial responsibility (Bouter, 2018), and it has to be done at an individual and a collective level. This can be done by promoting the organisation of courses and workshops, but also by responsible supervision practices (Resnik, 2012). Responsible supervision is essential for the promotion of RI, both individually and collectively, as supervisors act as role models for research practise and collaboration (Embassy editorial team et al., 2021). Research supervisors are responsible for engaging in the discussion on RI, with their supervisees (e.g. PhD candidates) and for enhancing the RI climate (ALLEA, 2017; Bell, 2015; Forsberg et al., 2018; Mejlgaard et al., 2020). Supervisors are not only responsible for transferring technical skills related to the research process and career advice, but they play also a pivotal role in fostering RI awareness and modelling fundamental attitude toward responsible research practices (Bell, 2015; Whitbeck, 2001). Besides instructing directly about good research practices and boundaries about what can be done or not, supervisors are exemplars for their doctoral candidates (Bell, 2015; Fisher, Fried & Feldman, 2009; Resnick, 2012; Rose, 2003; Weil, 2001). By acting as role models and showing specific virtues and characteristics, supervisors can influence and shape the behaviour and attitude of their PhD candidates (Abedin et al., 2012; Antes, Mart & DuBois, 2016; Gray & Jordan, 2012; Löfström et al., 2015; Rose, 2003). However, supervisors’ characteristics, practices, attitudes and behaviour can influence on supervisees’ research practices and understanding in terms of RI (Moncur, 2013; Muthanna & Alduais, 2021; Tijdink et al., 2016). It is not yet clear how other type of supervision (e.g. administrative, career, technological and supportive) other than ones RI-related, can influence PhD candidates’ behaviour and the entire research climate. Moreover, to our knowledge, empirical literature on possible differences depending on gender, seniority or discipline of supervisors and other variables is lacking. Although responsible supervision practices are main responsibility of the research supervisors, research institutions are responsible to support, help and assist supervisors in promoting RI practices, enhancing the general RI climate and putting them into the condition to fulfil at the best their role (Bird, 2001; Hauer et al., 2005). Dedicated training for supervisors and recognition for their efforts by institutions seems to be some of the needed steps for supporting supervisors (Bird, 2001; Kornfeld, 2012; Ripley et al., 2012; Titus & Ballou, 2014; Rose, 2003). While there are different conceptual papers on the importance of research institutions in supporting supervisors and responsible supervision (Bird, 2001; Hauer et al., 2005), to our knowledge, qualitative studies addressing supervisors’ opinions on what institutions can do to support them are missing.

This study aims to fill the empirical knowledge gap on how responsible supervision can influence individual responsible research and improve RI climate directly and indirectly, and whether differences in gender, seniority, discipline and other variables can have an impact on supervision practice. In addition, we want to preliminary explore empirically supervisors’ perspectives concerning the role of institutions in supporting responsible supervision. To this end, we conducted an interview study involving European research supervisors on these topics. Methods

We employed a qualitative methodology and organised a set of in-depth semi-structured interviews. We developed an interview guideline (Supplement 1-interview guideline) based on preliminary work on the topic (Pizzolato & Dierickx, 2022a). The interview guide has been created by using two main research questions. The first one is about exploring how supervisors’ behaviour and different practices can influence supervisees’ research practices and related behaviour. The second one is about how research institutions can support supervisors and responsible supervision to help them to promote RI and responsible research. During the interview, participants were also asked to rank pre-selected explicit and implicit supervision practices. These practices are described in the theoretical and empirical literature on the topic (Alfredo & Hart, 2011; Anderson et al., 2007; Brown & Treviño, 2014; Bukusi, Manabe & Zunt, 2019; Faden et al., 2002; Fisher, Fried & Feldman, 2009; Kornfeld, 2012; Titus & Ballou, 2014; Wright, Titus and Cornelison, 2008). Participants were also asked to indicate which virtues they think are important to act as a good supervisor and a role model, by choosing among virtues highlighted in the European Code of Conduct and literature (ALLEA, 2017; Character traits: Scientific virtue, 2016; Tomić, Buljan & Marušić, 2022). We followed the Consolidated Criteria for Reporting Qualitative Research guidelines (COREQ) to report the results of the study (Tong, Sainsbury & Craig, 2007). Not all the 32 criteria are met, namely experience and training of the research team, presence of nonparticipants, repeated interviews, and participant checking. The methods section is divided as follows: recruitment, data collection, data analysis and ethics.

Recruitments

To get insights regarding supervisors’ perceptions of their role as RI trainers and role models, we searched for supervisors with different characteristics. The potential interviewees were selected based on the following criteria: gender, country workplace, academic domain, seniority and whether they hold/held a European research grant(s) (e.g. European Research Council (ERC) grants, Coordinators of European Union (EU)-funded projects). We started the recruitment by asking for potential interested supervisors within our research network. Following a random sampling, we adopted a purposive sampling to identify and select supervisors with specific characteristics (e.g., gender, country workplace, academic domain, seniority and whether they hold/held a European research grant(s)). We invited them by email and by using an information letter describing the aim of the study (Supplement 2- Information letter). However, not to anticipate the real aim of the study and drive interviewees’ answers before starting the interview, we titled and described the study differently. We described the study as we were exploring supervisors’ role in the development of their PhD candidates, without mentioning the topic of RI. Supervisors interested in participating were asked to sign the informed consent and to fill in a brief demographic questionnaire (Supplement 3- Demographic questionnaire). We asked participants to choose among three academic fields, namely social sciences and humanities, life sciences, and physics and engineering. To avoid overlaps due to the multidisciplinary nature of some research, we decided to use the division used by the European Research Council (ERC- https://ec.europa. eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2022/wp\_horizonerc- 2022\_en.pdf). To assess the seniority of participants, we asked them the number of PhD candidates they have supervised, and the number of PhD candidates and post-doctoral researchers they are currently supervising (at the time of the interview). For the benefit of clarity and the writing process, we refer to supervisors who have supervised or are supervising less than ten PhD candidates in total as juniors and the ones more experienced as seniors.

Data Collection

We conducted 22 in-depth semi-structured interviews between October 2021 and January 2022. Due to the Covid-19 crisis, we carried out all the interviews online using Microsoft Teams. After 10 interviews, we adjusted the interview guidelines by adding one more question at the end of the interview (Supplement 1a- Modified interview guideline). During the first 10 interviews, some supervisors start discussing briefly the differences between being a good researcher and a good supervisor. Therefore, although quite late during the study, we decided to ask explicitly supervisors their opinion on the correlation between being a good researcher and a good supervisor. The interviews lasted on average one hour. After transcribing the interviews, no member checks were conducted.

Data Analysis

The data analysis was based on the thematic analysis approach and organised in two steps (Braun & Clarke, 2006; Vandemeulebroucke et al., 2019). Both authors participated in the analysis and coding of the interviews. In the first step, the transcripts were analysed by following the interview guideline and using a thematic inductive approach (Boyatzis, 1998, Vears and Gillam, 2022). This first step was performed alongside the interview process. The data were broadly coded with different colours to identify the main overarching themes. In the second step, the transcripts were re-analysed by using a deductive approach and narrowly coded to identify specific sub-themes (Crabtree & Miller, 1992). All the analysis process was carried out by using QRS NVivo 12 (NVivo qualitative data analysis software; QSR International Pty Ltd. Version 12).

Ethics

The interviews were performed after having obtained ethics approval by the Social and Societal Ethics Committee (SMEC) of KU Leuven under file number G-2021-3922. All the interviews were video-recorded and verbatim transcribed. After transcription, the video recordings were destroyed and the transcripts will be kept in a secure shared KU Leuven J-drive, accessible only to the researchers involved in the study.

Results

We performed 22 interviews with supervisors based in 16 countries (Fig. 1- Country distribution and Table 1- Demographics). During the interviews, supervisors discussed seven different themes related to their role and practices, namely supervisors’ role, supervisors’ responsibility level, supervisors’ practices concerning RI, supervision style, explicit practices, implicit practices and virtues expressed in role modelling. In these themes, the role of supervisors in promoting RI and how supervision practices can influence supervisees’ research practices are central. The results describe attitudes, behaviour and practices that can influence directly and indirectly RI at the individual and the collective level. Moreover, supervisors discussed three other themes related to the responsibilities of institutions in supporting supervision practices, namely training, institutional support and good supervision practices as criteria for assessment and promotion. An extra theme was discussed about the relation between being a good researcher and a good supervisor. Interesting quotes for each theme are collected in Table 2. In the results section, we highlight supervision-related differences depending on gender, discipline, seniority and country. However, there were no supervision-related differences concerning whether supervisors hold/held a European research grant. Supervisors’ Role

Most supervisors agreed on their main role being supervising PhD candidates as individuals and the development of their competencies. Supervisors should help, support and encourage PhD candidates in becoming intellectually independent as researchers during what was defined as “vocational training” (P12). Supervisors should make sure to train PhD candidates in strengthening their communication skills and their ability to work in teams. The latter was brought up mainly by supervisors working in the life sciences, engineering and physics. As highlighted by one supervisor, their main role is “to train the next generation of researchers” and “to train them how to navigate the academic landscape” (P20). A few senior supervisors highlighted how they should act on the development of a relationship with PhD candidates “based on trust and mutual respect” (P 5). So it’s really about finding the balance between guiding and letting someone develop himself (the PhD candidate). It’s a question of trust, and it’s not the same strategy for every candidate (P11). Another emphasised role by some female supervisors is being of “emotional support” (P18). Supervisors should emotionally take care of their PhD candidates by supporting them in stressful periods and in looking after their “mental health” (P21). Moreover, knowing whether something is happening within the personal sphere of the PhD candidates is important because can influence the PhD work. A few supervisors also mentioned their role in supporting PhD candidates with administrative formalities, role modelling regarding the work-life balance and with their future career steps. While discussing their different roles, some supervisors working in the field of engineering discussed the necessity to be good at recruiting PhD candidates to maximize the chances of success in both directions and to enable the completion of the project.

Concerning RI

Supervisors discussed specifically their role about RI. Some described their role as RI trainers as something intrinsic to their duty as supervisors. For them, transmitting RI competencies and attitude toward responsible research practices is part of guiding PhD candidates to becoming scientists and understanding their responsibilities.

Without integrity, there is no possible way to solve our scientific problem. I think this is associated with the idea of doing science (P1).

Some supervisors – working in countries where the topic of RI is more established – emphasised their role in providing direct and explicit instructions to their PhD candidates. The supervisor should communicate effectively as regards the standards, rules and procedures, including the legal ones and some customary rules of academy integrity and academic honesty (P5). Other participants emphasised their role in acting as a “good and responsible role model” (P5), in doing science but also in respecting colleagues. Some others also discussed their responsibility to provide PhD candidates with a responsible learning environment. Some supervisors – within the life sciences and engineering - emphasised their responsibility to make sure that all team members have a “common understanding of what the whole research process is” (P 1) in terms of responsible research practices. Moreover, a few reported also feeling responsible “to lower the stress and to take responsibility for their (PhD candidates) mistakes” (P 21). In relation to RI, supervisors discussed also their role when acting as co-supervisors. In terms of role modelling, they reported no differences in their behaviour. As highlighted by one participant, “it is about being professional” (P 2). In terms of providing explicit practices, their role depends on the level of RI skills and expertise of the main supervisor or other co-supervisors. Some participants – especially those working within the life sciences and engineering - emphasised the role of the supervisory team, where tasks, also related to RI, can be allocated to different supervisors depending on their level of experience and expertise. Supervisors also discussed their role as RI trainers when supervising post-doctoral researchers. Similarly, supervisors reported no differences in terms of role modelling. In terms of explicit practices, post-doctoral researchers are expected to be more independent in dealing with all ethical and integrity issues; therefore there is no need for supervisors to provide explicit training. However, a few supervisors expressed some concerns about having higher expectations from post-doctoral researchers. Post-doctoral researchers might fail in showing the same RI standards, especially when coming from other institutions or research groups.

Supervisors’ Responsibility Level

In responding to the question “how much do you think supervisors are responsible for your PhD candidates?”, supervisors reported different points of view. First, a few junior supervisors think that they are 100% responsible for the work, behaviour and research outcomes of their PhD candidates. They highlighted their responsibility “for reviewing everything” (P 14), since the research outcomes are the result of a “true collaboration with the PhD candidate” (P 14). Second, many supervisors – mainly seniors from the life sciences and engineering - reported feeling responsible for the technical-scientific part. However, since checking everything has been reported as an impossible task, “it is also a matter of trust” (P 15). Finally, a few supervisors – from the humanities and social sciences - reported feeling responsible only as a role model and for providing feedback if requested by the PhD candidate.

Supervisors’ Practices in Relation to RI The main activity described by most supervisors is the organisation of regular meetings, both within the research team and alone with the PhD candidate. However, there was no consensus about the frequency of the meetings. For some supervisors, regular team and one-on-one meetings must be organised weekly (e.g. life sciences, engineering and physics), for some bi-weekly, and for some others monthly or if needed under the request of the PhD candidate (e.g. social sciences and humanities). These meetings are usually focusing on the scientific-experimental part of the research, discussing RI- related issues only if needed. Some supervisors described these meetings as “informal training moments” (P 9) and as an opportunity to transform “possible mistakes into a learning moment” (P 21). Depending on the academic fields, discussions during the meetings focus on researchrelated issues, checking that the research is done properly (also from the technical/equipment point of view), reviewing data, data management, data analysis and how to report and properly communicate a study.

Supervision Style

All supervisors agreed that there is no a standardized way to supervise PhD candidates, but it depends on different variables related to experience, maturity, background and personal character traits of both, the supervisor and the supervisee. The supervision style is very much dependent on the period and on a specific situation during the PhD trajectory. However, one supervisor reported to have a common strategy for all PhD candidates concerning putting them “under the right level of pressure and stress” (P 17). According to the supervisors, the supervision style is depending on the learning curve of PhD candidates, but also on the learning curve of supervisors themselves. Mentoring is also how to learn to respond to students’ different needs (P 7) …after 30 years of mentoring, I feel I am learning every day. (P 1) The supervision style also depends on PhD candidates’ motivation and level of confidence. However, for a few supervisors being overconfident might be problematic, and that “students too keen to self-promote themselves have to be flagged” (P 15). A few senior supervisors highlighted that their style of supervising is also based and “dependent on the level of established trust” (P 17) and “common motivation” (P 5).

Explicit Practices

Supervisors were asked to rank the three most important explicit practices among the five pre-selected (Fig. 2 – Explicit practices). Most supervisors defined the five practices – or some of them depending on the academic field - as equally and extremely important. One participant reported not being involved in any of these practices, defining him/herself as “not a kindergarten teacher or a cop trying to bring candidates into lines” (P 2). Organising regular meetings has been ranked as the most important explicit practice, as a “standard way of mentoring” (P 19). This practice, together with reviewing data, is extremely valuable for supervisors working in the life sciences and, physics and engineering.

Implicit Practices

Supervisors were asked to rank also the three most important implicit practices among the six reported in Fig. 3.

It is much more difficult because they are all kind of equally important (P 12)

Supervisors ranked role modelling as the most important implicit practice to foster RI. The involvement of PhD candidates within the decision-making process has been reported to be also important since supervisors can explain better to PhD candidates why some decisions are taken and others not. According to some junior supervisors, also redefining the concept of failure and monitoring the level of stress of their PhD candidates are important to foster RI practices. In discussing these practices, one junior supervisor (physics and engineering) added to the proposed list one extra implicit practice. This practice is related to being as eco-friendly as possible in doing research and in transmitting this value to PhD candidates (e.g. do not print if not needed, lab-procedure on how to dispose of specific substances). Another junior female supervisor discussed the importance of paying attention to diversity-related issues (e.g. gender, religion and ethnicity).

Virtues Expressed in Role Modelling

Supervisors were asked to choose among 27 different virtues (Supplement 1-interview guideline), what they think are the 10 most important ones (without ranking them) that they have to express and show when acting as good role models. Honesty, respectfulness and availability are the virtues mentioned more often by supervisors (Fig. 4- Virtues expressed in role modelling).

Institutional Training for Supervisors

In the second half of the interview, supervisors were asked their opinions on the role of institutions in providing training on supervision practices and on RI. Supervisors expressed general appreciation for these institutional trainings and emphasised the need for institutions to take responsibility about this important subject. Institutions have to take responsibility for training their supervisors (P 1) The importance for research institutions to offer RI training for supervisors was clarified by one of the interviewees, by explicitly say that there are not enough supervisors trained in terms of RI. What I got in terms of research integrity, I got it from my students (P 11) Regarding the mandatory nature of both trainings, most of the supervisors were in favour to make them mandatory, for some also as “a contractual obligation” (P 8). Although considering mandatory trainings important, one participant expressed his/her perplexity about their implementation. In his/her opinion, “the people that make these decisions are professors as well and usually do not want mandatory things” (P 9). A few supervisors expressed some concerns about imposing mandatory training, seeing them more as an administrative tickbox exercise. The main concern regarding providing training sessions about their organisation. Courses are often led by someone without experience in guiding PhD candidates (P 14). According to supervisors, trainings on supervision practices should focus on communication skills, leadership practices, and how to deal with PhD candidates’ stress and psychological well-being. According to them, rather than being only based on theory, supervisors should engage into discussions with peers about their experiences. Training focuses on RI should be more discipline-specific rather than addressing general topics.

Institutional Support

All the supervisors agreed on institutions being responsible to support supervisors and supervision practices. They suggested different ways in which institutions can support supervision. First, by promoting informal exchange of experiences among supervisors, by organising regular group discussions. Second, by requiring at least two supervisors for each PhD candidate and by organising PhD supervisory committees at least once a year. Third, by having regular monitoring of the supervision practices or by having senior supervisors mentor juniors. Forth, by requiring a “formal written agreement between the supervisor and the PhD candidate” (P 12). Fifth, by “providing more funding and administrative support” (P 19). Finally, by supporting “candidates’ well-being and not leaving everything on the shoulder of the supervisors” (P 5).

Mart & DuBois, 2016). Intellectual and behavioural supervision are directly related to RI and can directly influence supervisees’ behaviour and research attitudes towards responsible research. Managerial and relational supervision are indirectly related to RI and can indirectly influence supervisees’ research practices in terms of QRPs and misconduct. Failing to facilitate managerial issues related to the PhD trajectory and to promote a healthy supervisor/ PhD candidate relationship might put the PhD candidate under pressure and in search of shortcuts or more willing to engage in QRPs. This has been also discussed extensively in the literature, where it has been highlighted that reducing the PhD candidates’ stress and pressure, and increasing their mental well-being is beneficial to reducing QRPs (Moncur, 2013; Redman et al., 2006; Roberts, Kavussanu & Sprangue, 2001). However, it is unlikely that a single supervisor can have enough competencies, skills and time to deal with all the different activities that have been described. As reported in the literature, having multiple supervisors would allow PhD candidates to benefit from different approaches, skills and competencies, also in terms of filling possible gaps in promoting responsible research practices (Abedin et al., 2012; DeCastro et al., 2013; Haven et al., 2020; Pennanen, Heikkinen & Tynjӓlӓ, 2020). Moreover, supervisors can benefit from a sharing activity-related responsibility. In the interviews, it was made clear by the participants that even if not all supervisors can carry out all the related activities, the role model function must be a core element, also in co-supervision of students. Group mentoring would be also beneficial to PhD candidates in case one of the supervisors is not willing to engage with explicit activities aiming to foster RI (as reported by one of the interviewees). In a classic one-on-one supervision relationship, this inadequate supervision approach might be harmful to the PhD candidates’ well-being, their perception of RI practices and responsible supervision, and it can lead them to engage in research-related misbehaviour. While conducting the interviews, it became clear how supervision practices are not standard and depend on different variables. Although these variables are often related to the supervisee and its learning curve, attitudes, competencies, maturity and motivation, supervision practices depend also on supervisors’ characteristics, experience and willingness to engage in supervisory activities. Supervisors’ practices in relation to their role of promoting and facilitating good research practices and fostering RI standards have been described differently depending on seniority, country, academic discipline and academic disciplines. During the interviews, it was interesting to note that the EC grant record of the supervisors has no impact on the supervisors’ view of general supervision and RI practises. The only difference is in relation to open science practises. Supervisors who have an EU grant seem to be more willing to engage in open science and open publications than other colleagues. This may be a consequence of the fact that the EU requires mandatory open access publications for its grantees. In relation to role modelling, supervisors consider it crucial and fundamental in order to show supervisees a model they can imitate and be inspired by. Senior supervisors see themselves as RI trainers as something implicit to their role. The development of the scientific competencies of the PhD candidates can not be separated from the development of RI competencies and the internalization of professional values and RI high standards. This is also emphasised in the literature. “Professional supervision” benefits PhD candidates in becoming responsible researchers and future responsible supervisors (Abedin et al., 2012; Huybers, Greene & Rohr, 2020; Ripley et al., 2012; Straus, Chatur & Taylor, 2009). Senior supervisors do not feel the need to be in control of every step taken by the PhD candidates. Their control decrease depending on the learning curve of the super- visees. Senior supervisors value the creation and the continuation of a relationship based on trust and mutual respect. A relationship that Lee and colleagues described to be “for life” with a future well-supported colleague (Lee, Dennis & Cambpell, 2007). Juniors seem to be more willing to be involved in/in command of all the research phases and in all the steps of the research trajectory. Juniors seem to be more willing to pay attention to emerging issues such as the level of stress of their PhD candidates and address their fear of failing. As highlighted in literature, these practices might be useful to lower the pressure on PhD candidates and to avoid them to engage in QRPs or consider taking shortcuts (Redman et al., 2006; Roberts, Kavussanu & Sprangue, 2001). In addition, juniors seem to be more aware of their societal and environmental responsibilities. The importance of supervisors addressing these emerging RI topics has been already stressed in previous studies (Bouter et al., 2016; Pizzolato, Abdi & Dierickx, 2020; Watts et al., 2017). Moreover, as already stressed in literature, junior supervisors seem to feel more responsible for being exemplars concerning the work/life balance (Straus et al., 2013). Supervisors working in countries where national regulatory documents, and apparently training, concerning RI are present for far longer (Desmond & Dierickx, 2021), seem to be more willing, inclined and prepared to discuss the topic. The major willingness to discuss RI, rules, guidelines and engage in explicit practices of supervisors working in specific countries (e.g., Scandinavian countries) seems to reflect the fact that the topic is nowadays well-rooted into the research system. In these countries, research practices and RI go hand-in-hand and are not taught separately. Moreover, it seems that supervisors working in countries where RI training, usually for PhD candidates, are more present seem to be more willing to address the topic (Abdi et al., 2021). The majority of the supervisors value having regular meetings (in team or one-on-one) with their PhD candidates. However, frequent regular meetings seem to be a priority only within the life sciences and, physics and engineering. Explicit practices discussed during the interviews seems to be specific to the research team environment and academic field. As reported also by previous studies, the review, analysis and interpretation of raw data seem to be extremely important in lab-related environment (Antes et al., 2019; Rabatin et al., 2004; Titus & Ballou, 2014). Within the social sciences and humanities, supervisors appear less willing to engage in explicit practices and in socializing PhD candidates into RI. Differences in RI expectations among academic fields has been already reported in literature, where RI expectations regarding the social sciences and humanities were lower than for other academic fields (Haven et al., 2019; Wells et al., 2014). Female supervisors seem to be more willing to be of emotional support to their PhD candidates than their male colleagues. In discussing their role, female supervisors seemed to be personally and emotionally attached to their supervisees, not just in relation to their PhD trajectory, but also in relation to personal issues. Taking care of them and being of emotional/psychological support is perceived as part of their role. The importance of being of emotional support has been also highlighted by previous studies (Abedin et al., 2012, Anderson et al., 2007). Being of emotional support might lower the stress and pressure of PhD candidates and the possibility to engage in QRPs (Redman et al., 2006; Roberts, Kavussanu & Sprangue, 2001). On the opposite, knowing what is going on in the personal life of the PhD candidates is also important for male supervisors, but it seems to have more a practical scope. Preventing possible unintentional questionable practices by knowing that the PhD candidate might lose focus on his/her work or engaging due to a personal situation.

Although there is a correlation between specific supervision practices and supervisors’ characteristics (e.g. seniority, country, academic discipline, academic disciplines), the relationship that is created between the supervisor and the PhD candidate is unique and different from others. As any other relationship, it depends on uncountable variables. “Having the right chemistry” is important to foster a trust-based and healthy relationship (Jackson et al., 2003; Straus et al., 2013). As highlighted by the majority of the supervisors during the study, being a good researcher is just a prerequisite to being a good supervisor. Being a good and responsible researcher is crucial, but it differs from being able to transmit responsible practices and to follow the development of PhD candidates. Supervisors have to express a set of virtues that slightly differ from the ones important to being a good researcher. Honesty has been highlighted by many as the most important virtue. The same can be found in literature where honesty is defined crucial to be a good supervisor as well as a good researchers (Character traits: Scientific virtues. Nature, 2016; Pizzolato & Dierickx 2022a; Straus et al., 2013). Virtues identified as important to be a good researcher (e.g. perseverance, objectivity, humility of evidence and attentiveness) are replaced by virtues considered essential to be a good supervisor (e.g. respectfulness, availability, creativity and positivity) (Character traits: Scientific virtues. Nature, 2016). Respectfulness, availability, positivity were also highlighted by Lee and colleagues in their guide for mentors published on Nature (2007). In addition to different virtues, good supervisors must have specific competencies and characteristics that differ from the ones that a good researcher must have. Possessing the right virtues is just the starting point to act as a supervisor. Good supervisors must possess a set of competencies that complement the ones for being a good researcher. These competencies can be innate in some researchers, but they can also be taught and trained. The supervisors participating in the study clearly valued the organisation of specific institutional training on RI and supervision practices targeting supervisors. According to the literature, these trainings play an important role in fostering supervisors RI practices and ability to transmit these practices (Löfström et al., 2014; Ripley et al., 2012; Straus et al., 2013). Recently, Dutch supervisors participating in a pilot study described this typology of training as highly recommended and reported to be more skilled than before participating in the training (Haven et al., 2022). Moreover, supervisors might be involved in dedicated sessions led by some junior colleagues, who might be more prepared in addressing RI issues (Pizzolato & Dierickx, 2022b). Supervisors have made suggestions as to how institutions can support supervisory practices. Although this institutional support can only be seen in terms of promoting supervisory practices and supporting supervisors, it has a cascading effect on supervisees in terms of research practises and associated attitudes and behaviours. Creating informal group discussions would enable supervisors to exchange best practices but also tips and tricks on how to deal with specific situations (e.g., lowering supervisees’ stress and addressing their fear of failure). The creation of these peer-discussion groups has been also previously advocated within the literature (Di Benedetto et al., 2021; Gruber et al., 2020). Requiring at least two supervisors would benefit the PhD candidate but also the supervisors involved. PhD candidates, and supervisees in general, would benefit from different perspectives, not to have one world view of RI practices and the academic environment (Haven et al., 2020). Supervisors would benefit from sharing commitment and time in training the PhD candidates. Moreover, having more than one supervisor would probably lower possible tensions that a one-on-one relationship can create and that can lead to a stressful and unhealthy collaboration, and as a possible consequence, it can lead supervisees to engage in QRPs. Matching senior supervisors with junior supervisors might be beneficial to provide informal training sessions to complement institutional training on supervision practices. Besides what suggested, a few best practices are already reported within the literature (Lechner et al., 2020; Lerouge & Hol, 2020, Sawatzky & Enns, 2009). The supervisors participating welcomed the inclusion of supervision practices within the pool of criteria for assessment and promotion. As institutions assess academic performance in teaching, also supervision practices must be evaluated and considered. As already stressed in literature, formally recognizing the role of supervisors in training “the next generation of researchers” is something that institutions have consider (Kafedjiska et al., 2022). In line with the San Francisco Declaration on Research Assessment (DORA) and the Hong Kong principles (Moher et al., 2020, principle 5), institutions should not merely focus on journal metrics but also consider the larger contribution that researchers, in acting as supervisors, can give to the research environment. The main concern of the supervisors involved in the study was how to make this criterion objective and measurable. A qualitative and subjective bottom-up assessment by doctoral students and supervisees can help to get an idea of certain supervision practises. Although this assessment can only give half a picture of the real situation, multiple concerns about a particular supervisor may prompt institutions to monitor the situation closely.

Strengths and Limitations

To our knowledge, this is one of the few qualitative studies involving focusing on supervisors’ role as RI trainers (Pizzolato & Dierickx, 2022a). The study investigates supervisors’ perceptions concerning their RI practices and responsibilities across country, gender, academic domain and seniority. This study can serve as a preliminary attempt to highlight differences in supervisors’ perceptions, approaches and real-life practices as RI trainers and role models. Moreover, this explorative study can serve as an outline to organise further research on the matter. However, it is also fair to point out some limitations. The most important limitation of this study is the selection bias. At this stage, we are aware that it is not possible to generalise and consider these study results as universally valid. The sample size is probably too small to understand in depth all the possible nuances arising from the different inclusion criteria. Although we found differences between the three different academic domains, we are aware that there are also differences within each academic domain. The same reasoning can be applied to the differences between countries. Although we have included supervisors from 16 different European countries, not all countries are represented (e.g. Germany, Austria, the Baltic countries). Unlike the US, where RI is firmly rooted throughout the country, European countries have a different understanding, tradition, research system and policies/ laws regarding responsible conduct. In most cases, while there are clear differences between countries, in some countries there are also differences between institutions or even faculties. As for junior supervisors, while there are some commonalities, their approach to RI may have been influenced by too many variables, and the sample size available is too small to understand all possible nuances in depth. More in-depth studies of supervisory practices that focus on a specific criterion (e.g. gender, discipline, country and seniority) are needed to address these issues.

Conclusion

This study tries to give an overview on supervisors’ perspective concerning their role as RI trainer and role models, and how supervision practices can influence supervisees’ research practices. In addition, it tries to give voice to supervisors regarding what institutions can do to support them and supervision practices. Supervisors value their role as responsible exemplars, in fostering RI awareness and in transferring RI competencies. Although there are some commonalities concerning supervisors’ understanding of their role concerning role modelling, approach and real-life supervision practices can differ largely across discipline, gender, seniority and country. Although it is possible to outline different best practices (e.g. having regular meetings, discussing good research practices, monitoring PhD candidates’ stress), there is no a unique and standard way to be a good and responsible supervisor. A successful collaboration, transmission of professional values and RI competencies also depend on the presence of a human match between the supervisor and the PhD candidate. All supervisors agreed on institutions being responsible in supporting supervisors and supervision practices. This can be done by providing dedicated training sessions on RI and supervision practices. They provided also a few examples on how institutions can assist supervisors and promote responsible supervision practices (e.g. peer discussion groups, coupling senior and junior supervisors, providing administrative support). Although supervisors valued the inclusion of supervision practices within the criteria for promotion and assessment, they expressed concerns regarding making this criterion as objective and measurable as possible. This explorative study puts the basis to plan new and more specific studies. Further focused studies are needed to in-depth understand differences across gender, academic domain (also within the same domain), country and seniority. Moreover, studies focusing on making supervision practices measurable and objective are very much needed to start considering supervision on the same level of journal metrics.

R&D-based endogenous growth models, starting with Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), have demonstrated that technological progress relying on cumulative R&D expenditure accounts for much of output growth in the last century. Given that this proposition is mostly about industrial R&D, it raises an intriguing question about the contributions of academic research to this economic development process. In the OECD area, member countries have injected substantial funding into research conducted in academia every year (OECD, 2021). Many economists have attempted to explore the effect of academic research on our economy and society. So far, the results obtained are mixed. In particular, there are several papers that find evidence on the spillovers from academic research to industrial innovations (e.g. Cohen et al., 2002; Jaffe, 1989; Le et al., 2022). Studies by Mansfield (1991, 1998) indicate the important role of academic research in developing new products and processes, as well as in contributing to greater social benefits. Meanwhile, there are also studies that report a negative relationship between scientific research and innovative output produced by the industry (e.g. Gittelman and Kogut, 2003; Partha and David, 1994). Those that advocate academic research point out that, thanks to its widespread through publications in scientific journals, academic research knowledge resembles a public good and creates important scientific foundations for technological progress in the industry. Besides scientific papers channel, this knowledge can diffuse to industrial production in other ways, such as through the mobility of students and scientists (Zucker et al., 2002), R&D collaborations between universities and companies (Jaffe et al., 1993) or direct licensing from universities to private firms (Jensen and Thursby, 2001; Thursby and Kemp, 2002). This diffusion process can help enhance productivity (Le et al., 2022). In that respect, Berggren and Bjornskov (2022) point out that academic freedom is beneficial for long-term economic development because it allows creative and productivity-enhancing ideas to be developed and diffused to the industry without interference. Meanwhile, proponents of academic research argue that academic research is mostly aimed at recognition and promotion in academia and, hence, has little economic value (Dasgupta and David, 1994). This means that the issue on the role of academic research, especially academic research at the top or frontier level, in affecting economic performance continues to be a highly debatable topic. Against this background, the main goal of this paper is to shed further light on the debate about the role that frontier academic research plays in affecting total factor productivity (TFP). Our argument is that frontier academic research is not just limited within the aim of promoting the reputation of the authors, but more importantly, it paves the way for industrial innovations to flourish, which in turn improves technological development. Using a dataset that covers 18 OECD countries over the 2003– 2017 period, we find evidence that frontier academic research exerts an important effect on TFP. We find statistical support to the direct effect of frontier academic research on TFP. We also find significant evidence that frontier academic research affects TFP via the industrial R&D channel. As such, our empirical results confirm that industrial R&D is an important conduit through which frontier academic research induces TFP. In examining the conversion from research and innovation into technological progress, we extend our empirical analysis to consider the impact of institutional factors on these processes. In that respect, our purposes are twofold: (i) evaluating the impact of institutional factors on the degree to which frontier academic research is absorbed into industrial innovation; and (ii) assessing the extent to which industrial R&D is materialised into TFP. We focus on the economic aspect of institutional quality by using an index of economic freedom and different dimensions of this index. This index is composed by the Fraser Institute in an attempt to measure the degree to which an economic and political system allows people to specialise and trade.1 We find that countries with a relatively higher level of economic freedom tend to reap more benefits from frontier academic research. However, different dimensions of economic freedom alter the way countries enjoy the benefit from frontier academic research differently. On the frontier academic research and industrial R&D nexus, there is little evidence that institutional quality enhances the efficiency of the process that turns frontier academic research into industrial innovations. Regarding the effect of industrial innovations on TFP, we find that the effect is strengthened with an increase in the degree of economic freedom. Nevertheless, the corresponding moderating effect of different components of this index is mixed. In particular, while the moderation is positive and significant with government size and the conduct of monetary policy, it is negative and significant with the protection level of private property rights. As for free trade and limited regulation measures, the moderating effects are insignificant. Our paper contributes to two different strands of literature. First, it contributes to the literature examining the economic impact of academic research. These studies indicate several ways through which academic research can induce technological change such as education and training (e.g. Bekkers and Freitas, 2008; Jones and de Zubielqui, 2017; Rosenberg and Nelson, 1994), university– firm research collaborations (e.g. Acs et al., 1992; Audretsch et al., 2012; Cohen et al., 2002; Faulkner and Senker, 1994; Lai, 2011; Thursby and Kemp, 2002; Wirsich et al., 2016) and academic publications (e.g. Audretsch, 2013; Lundberg, 2017; Zucker et al., 1998). Second, this paper contributes to the literature considering the impact of institutional factors on economic development. According to Guellec and van Pottelsberghe de la Potterie (2004), institutional settings affect the contribution of knowledge to productivity. Tebaldi and Elmslie (2013) and Krammer (2015) point out that the effects of institutional quality on R&D activities and economic performance vary across countries and institutional elements. Taking Chinese economy as an example for a transitional economy, a study by Hou et al. (2021) shows that poor institutional quality such as administrative interventions and limited degrees of freedom can adversely affect the universities’ tendency to explore technology transfer channels and conduct academic entrepreneurship actively. Meanwhile, better institutions with a more business-friendly environment and lower barriers to trade and investment may amplify the positive impact of R&D spending on multi-factor productivity (Égert, 2016). Similarly, strong judicial accountability is an important condition for academic freedom to contribute to economic growth (Berggren and Bjornskov, 2022). Although there is a growing body of literature confirming institutions as deep determinants of growth (e.g. Acemoglu et al., 2001; Rodrik et al., 2004), a much smaller number of papers attempt to investigate the roles of institutions within an R&D-based context. Notable exceptions include Coe et al. (2009) and Krammer (2015) with an assessment of how institutions affect R&D spillovers. They find that a higher quality of institutions is associated with larger international R&D spillovers. To a broader extent, given that a large proportion of academic R&D comes under the form of public investment, this paper is also related to the literature assessing the productivity effect of public R&D (e.g. Guellec and van Pottelsberghe de la Potterie, 2004; Soete et al., 2020a, 2020b). Studies in this strand of literature point out the important complementarity between public and private R&D. Our paper proceeds as follows. In section 2, we describe data collection as well as the way we construct data series for the empirical investigation. We also present summary statistics of the data used. In section 3, we discuss our empirical strategy. We conduct some preliminary data analysis using panel cointegration in section 4. In section 5, we present estimation results on the nexus among frontier academic research, industrial R&D and TFP. While section 6 is devoted to the examination of the moderating effects of institutions, section 7 discusses robustness check results.We end the paper with some concluding remarks in section 8.

2. Data construction and summary statistics

2.1. The measures of frontier academic research

Our measure of national frontier academic knowledge is computed based on the data on research capability of Top 500 universities worldwide known as Academic Ranking ofWorld Universities (ARWU) published by Shanghai Jiaotong University (since 2003).2 Research strengths of universities are assessed and scored according to six indicators: the number of alumni awarded Nobel prizes and Fields medals, the number of staff awarded Nobel prizes and Fields medals, the number of highly cited researchers, the number of papers published in Nature and Science,3 the number of papers indexed in Science Citation Index (Expanded) and Social Science Citation Index, and the per capita academic performance of these indicators. Although more than 1,000 universities are surveyed each year, only the rankings and scores accompanying the rankings of the Top 500 universities are reported. Among these indicators, it is arguable that for academic knowledge, publications are the most influential factors on technological development. On this ground, we focus on the scores on indexed publications. 4 If we denote Pubkjt as publication scores for ARWU-listed university k that is located in country j and in year t, by aggregating the scores of all listed universities for each country over the same year we obtain the national stock of frontier academic publications as follows:

Pubjt = k Pubkjt .

(3) This variable is considered to contain a substantial level of frontier academic knowledge as it measures academic research conducted at leading academic institutions in the world.

2.2. The measure of industrial R&D

To construct the measure of industrial R&D, we follow the steps of calculating total R&D capital stocks suggested by Coe and Helpman (1995) and used by subsequent papers (e.g. Bayoumi et al., 1999; Engelbrecht, 1997; Le, 2008, 2010, 2022; Le and Bodman, 2011) in the R&D-based growth literature. After getting data on nominal industrial R&D expenditure from OECD Statistical Database, we deflate it by an R&D price index to generate real R&D expenditure, RD, before moving on to calculate the R&D capital stock, SD, such that SDit = (1 − d)SDit−1 + RDit−1 (the depreciation rate d is chosen to be 5%). The stock at the beginning of the period is specified as Si0 = RDi0/d + g where g is the annual average growth rate from 2003 to 2017 generated within the data.

2.3. The measures of institutional quality To characterise institutions, we collect data on the degree of economic freedom of countries on a 0–10 scale composed by the Fraser Institute. This measure of institutions has been used extensively for conducting institutional analysis (e.g. Berggren and Nilsson, 2021; Bergh, 2020; Dean and Geloso, 2022; Graafland, 2020; Kufenko and Geloso, 2021; Moellman and Tarabar, 2022). The index is calculated based on 41 indicators that can be categorised into five different areas: (i) government size (government spending, tax burden and fiscal health); (ii) private property rights (property rights, government integrity and judicial effectiveness); (iii) sound money (monetary policy and control of inflation); (iv) free trade (tariffs, quotas, financial capital controls and international travel); and (v) limited regulations (business freedom, labour freedom and monetary freedom). We will first consider the effect of institutions through the use of the aggregate index before examining each area of specialisation in greater details as each aspect of institutions is likely to affect innovation and technological development path in a different way.

2.4. The measure of total factor productivity We use TFP to proxy for technological improvement of a country because this is the factor that explains cross-country differences in GDP per capita over the last century (Caseli, 2005; Hall and Jones, 1999). To calculate this variable, we start with the Cobb–Douglas production function: Yit = FitKg itL1−g it where Yit is the value-added, Fit is TFP, Kit is capital stock and Lit is employment in the business sector. Data for Y, K and L are extracted from the World Development Indicator Database provided by the World Bank. From this data series, we regress the value-added on capital stock (constructed from data on capital formation) and employment to get γ. After that, we calculate TFP for each country in the sample using the following formula: log (Fit) = log (Yit) − γlog (Kit) − (1 − γ)log (Lit). The computed series is then converted to an index in which the value in 2011 is set equal to one.

2.5. Other variables

Data on human capital stock are extracted from the Penn World Table (version 9.0). This variable will be used as a control variable in all of our regressions. Available data include an index of human capital per person, which is computed based on information on average years of schooling as per Barro and Lee (2013) as well as that on return to education as per Psacharopoulos (1994).7 Another control variable is trade openness. According to Miller and Upadhyay (2000), larger trade facilitates the adoption of more efficient techniques of production leading to faster growth of TFP. We collect data on trade openness (i.e. sum of exports and imports) as percentage of GDP from the World Bank. We establish our sample based on the conditions that the selected countries must have had a large number of universities listed in the ARWU over the whole 2003–2017 period. They also need to have recorded data on industrial R&D expenditure. In the end, we obtain a balanced panel data set that covers 18 OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Netherlands, Norway, Spain, Sweden, Switzerland, the United Kingdom and the United States.

2.6. Summary statistics

In Table 1, we provide a summary of key data series for 18 OECD countries over the 2003–2017 period. It can be seen that TFP slightly decreased in most countries except for Finland, Ireland, Netherlands and Spain where there was a small increase and Norway and the US with roughly no change. Meanwhile, industrial R&D capital stock increased substantially in all countries with highest increment recorded in Denmark, followed by Ireland and Australia. By contrast, Japan and Italy experienced the slowest expansion of industrial R&D. Changes in frontier academic publication scores are not as dramatic as industrial R&D. While Australia experienced the fastest expansion of more than twofold, Japan and Italy faced a sharp fall of this research stock. Other countries are divided, either enjoying a slight increase in scores, like Belgium or Denmark, or seeing a small downturn, like Canada or the UK. As for human capital stock, all countries had moderate improvements. These improvements exhibit a somewhat homogeneous pattern across countries. Italy and Spain enjoyed the largest increases while Australia and Germany only had modest rises. Turning to trade openness as a percentage of GDP, all countries experienced substantial improvements over the time horizon except for Canada with a small decrease. The biggest winner is Japan, followed by Ireland and Germany. Regarding variables capturing institutional quality, change in the aggregate index of economic freedom, alongside with that for the five component areas, is reported for all countries in the sample. No single countries advanced in all areas of consideration although they made some improvements overall. Countries that enjoyed biggest increments overall include Australia, Germany and Norway. Meanwhile, the UK, Spain and the US are those experiencing most deterioration in the aggregate index.

3. The empirical model

Our interest is to investigate the way through which frontier academic knowledge contributes to technological development of a country and how institutional quality affects this process. Because academic research is often considered to provide basic scientific results that lay the foundation for technology to flourish, we hypothesise that frontier academic knowledge exerts its effect on technological improvement through different channels. In the direct channel, frontier academic research stimulates technological progress by directly providing knowledge that helps improve management efficiency and production methods. Meanwhile in the indirect channel, the contribution of frontier academic knowledge is seen through its stimulation to industrial R&D and innovations, which, in turn, enhances technological change.8 In both processes, institutions can matter as they may create either barriers or opportunities for research (either academic or industrial) to materialise into technological improvement. Figure 1 below provides a graphical representation of the channels. Path A in the figure captures the total effect of frontier academic knowledge on TFP. To explore this path, we run the following regression:

log (Fit) = ai +a1 log (Pubi,t−2) +a2INSi,t−1 +a3INSi,t−1 × log (Pubi,t−2) +dXi,t−1 +gt + 1i,t , (4)

where F is the TFP index, Pub is the frontier academic research capital stock measured in terms of publication scores, INS is an indicator of institutional quality, X is the vector of control variables such as stock of human capital and trade openness, αi is a country fixed effect that picks up effects of time-invariant factors on technological progress such as culture or climate, γt is a time fixed effect that absorbs time-varying characteristics such as macroeconomic shocks, and 1 is an error term. Given that frontier academic research is not performed by the industry, a longer delay is expected before it can affect technological level so it enters the equation with a 2-year lag.9 Path B implies that frontier academic knowledge may induce industrial R&D with the moderating effect of institutions. To examine this possibility, we put forward a regression equation as follows:

log (SDi,t−1) = ai +a4 log (Pubi,t−2) +a5INSi,t−1 +a6INSi,t−1 × log (Pubi,t−2) +gt + 1i,t , (5)

where SD is the measure of industrial R&D capital stock. Path C in the figure helps reveal the impact of industrial R&D on TFP. This path can be explored by the regression equation below:

log (Fit) = ai +b1 log (SDi,t−1) +b2INSi,t−1 +b3INSi,t−1 × log (SDi,t−1) +dXi,t−1 +gt + 1i,t . (6)

The direct effect of frontier academic knowledge on TFP is captured by path D with the following regression equation:

log (Fit) = ai +b4 log (SDi,t−1) +a7 log (Pubi,t−2) +a8INSi,t−1 +b5INSi,t−1 × log (SDi,t−1) +a8INSi,t−1 × log (Pubi,t−2) +dXi,t−1 +gt + 1i,t . (7)

4. A preliminary analysis using panel cointegration

In this paper, we will conduct our estimation using panel cointegration methods. Since inception, panel cointegration has been used widely in the R&D-based growth literature to establish long-run relationship between non-stationary variables (Coe et al., 2009; Coe and Helpman, 1995; Le, 2010, 2012; Le et al., 2022). For that purpose, we first implement panel unit root tests on the variables. At 10% level of significance, Hadri’s (2000) and Im et al.’s (2003) tests, reported in Table 2, indicate the overall non-stationarity on the majority of variables. There are a few exceptions including trade openness as a percentage of GDP, aggregate economic freedom index and some of its sub-components such as government size, private property rights and free trade. For these variables, test results indicate non-stationarity under Hadri’s (2000) test but stationarity under Im et al.’s (2003) test. In making our own judgement, we tend to rely more on the outcome from Hadri’s test given our purpose of proving the variables to be non-stationarity. This is because Hadri’s test has the null hypothesis of stationarity on the variable while Im et al.’s test projects the existence of an individual unit root process in the null hypothesis instead.

We next examine if the combination of the time series exhibits a co-integrating relationship. This is the statistical requirement for having meaningful estimations (i.e. regression results are not spurious). Reported results at 10% level of significance on two panel cointegration tests put forward by Pedroni (1999) in Table 3 reveal that this requirement is satisfied for most regressions as both tests confirm the existence of cointegration among variables. In a few cases involving some institutional variables, while the panel augmented Dickey-Fuller (ADF)-statistics indicates cointegration, the group ADF-statistics do not confirm such a relationship. To make a decision, we are inclined towards using the outcome of the panel ADF-statistics as the corresponding test pools the statistics along the within-dimension rather than averaging the results of individual country test statistics as the group ADF-statistics do. According to Kao et al. (1999) and Tsionas (2019), in dealing with co-integrated panels, Ordinary Least Squares regression (OLS) results may be subject to a second-order asymptotic bias due to the endogeneity problem that is caused by the potential reverse causality between R&D variables and TFP. Following Kao and Chiang (2000) and also to save space, in what follows, we will only present regression results conducted using the dynamic OLS (DOLS) method.10 The advantage of this method lies in its superior small sample properties, which are more suited with our sample. To preserve the number of observations, we choose one lead and one lag of the cointegrating regressors for all of our regressions.11

5. Frontier academic research, industrial R&D and technological development

In Table 4, we report DOLS results for regression equations (4)–(7) in which frontier academic publication scores are used to represent frontier academic knowledge. Note that in running these regressions, we withhold from considering the role of institutions in order to solely focus on the relationship between frontier academic research, industrial R&D and technological progress. Except for regression equation (5), all equations include human capital and trade openness-GDP ratio as control variables. They also include unreported country-specific fixed effects (FEs) to control for factors that affect TFP but do not vary little with time such as geographical and climate conditions. Additionally, timespecific FEs are used to take account of common and time-varying factors that potentially affect TFP across countries such as economic crisis or other macroeconomic shocks. For equation (4) on the total effect of publications on TFP, column (4.1) indicates that frontier academic research has a positive and significant overall impact on TFP as captured by a positive and mostly significant coefficient of log (Pub). This result is in line with those previously obtained by Le and Tang (2015) and Le et al. (2022). Estimation result for equation (6) on the total impact of industrial R&D on TFP is given in column (4.2). It can be seen that industrial R&D strongly enhances technological improvement as the coefficient on log (SD) is positive and highly significant. This adds more evidence to the one previously established in the literature such as Coe and Helpman (1995) and Le et al. (2022). Regarding the direct effect of frontier academic research on TFP as per equation (7), column (4.3) indicates no significant evidence for the existence of such an effect. While the coefficient of industrial R&D variable continues to be positive, it is insignificant. This means that in the presence of industrial R&D, the direct effect of frontier academic research on TFP is not clear. As for regression equation (5), our estimation result is provided in column (4.4). In can be seen that frontier academic knowledge induces more industrial R&D as evidenced by a positive and significant coefficient of log (Pub). This implies a complementarity between frontier academic research and private sector R&D. This ‘crowding-in’ effect, as explained by Cassiman and Veugelers (2002), is partly due to the potential attraction to internationally mobile R&D. This includes factors related to prospects of highquality collaboration, recruitment opportunities and technological transfer infrastructure.

6. Institutions as a moderating factor

In this section, we examine if the estimated coefficients on the nexus between frontier academic research, industrial R&D and TFP vary due to the introduction of institutional variables. While the views on institutions are many, ranging from political to cultural and economic aspects (La Porta et al., 1999), we focus only on the economic aspect of institutions. In particular, we pay attention to a broad-based index of economic freedom and its sub-components including the effectiveness of fiscal policy, the strength of private property rights, the effectiveness of monetary policy, the freedom of trade and the degree of regulatory control. Each of these variables could potentially affect the extent and direction in which frontier academic research and industrial R&D affect TFP. Estimation results for the aggregate index of economic freedom are reported in Table 5. We start with simple regressions that examine the direct effect of economic freedom on TFP while taking into account the influence of frontier academic knowledge (in column (5.1)) and industrial R&D (in column (5.3)). We then test the indirect effect of economic freedom on TFP by including an interaction term of this indicator with each of the technological knowledge variables in regressions (5.2) and (5.4) respectively. All of these regressions include human capital stock and trade openness as control variables. It can be seen that economic freedom exerts little direct effect on TFP since its estimated coefficient is mostly insignificant. Meanwhile, there is evidence that economic freedom indirectly affects TFP. This is because the estimated coefficients for the interaction terms of economic freedom with each of the technological knowledge variables are both positive and statistically significant. This means that greater economic freedom enhances the impact of technological knowledge, either academic or industrial, on productivity of countries. In columns (5.5) and (5.6), we investigate the effects of economic freedom on industrial R&D investment. While the direct effect is negative and statistically significant, the indirect effect (i.e. via frontier academic knowledge channel) is negative but insignificant. The intuition for the above results is as follows. The economic freedom index captures a wide range of aspects related to doing business. It includes pro-business market reforms that make it easier for investors to start and run a business. While this stimulates entrepreneurship, it raises the level of competition among the firms. To thrive in the market, some firms may choose to imitate others’ technology to improve their productivity at low cost instead of developing their own technology. This process will somehow chip away the potential monopoly profit earned by a future successful innovator. In response to this threat, firms may consider cutting down their R&D investment. By construction, economic freedom also refers to the set of institutional standards such as rule of law and open mark regulations. An increase in this score implies changes that enhance efficient allocation of resources. Because firms and countries in our sample have already been enjoying a good environment that stimulates innovative capabilities,12 they will gain little where there is a lower level of regulation or an enhancement of flexibility. The result that there is a positive moderation of institutional variables on the relationship between frontier academic knowledge and TFP can be explained as follows. Because academic research resembles a public good, it is generally accessible by the public. However, the majority of academic research belongs to basic and theoretical science meaning that it is not ready to make a real impact on the economy in its original form. An improvement in economic freedom encourages entrepreneurship, which in turn leads to more application of scientific results into industrial innovation. This can be done in the form of university–industry R&D collaborations or education and training. As a result, firms in countries of higher economic freedom will reap more benefits from academic research. In Table 6, we report results obtained from performing a similar test, however, using a component of the economic freedom index, namely government size. The results reveal a negative and statistical significant direct effect of government size on TFP. Nevertheless, the indirect effect is positive and significant. In addition, government size also influences industrial R&D but it does not affect the way that frontier academic research impacts industrial R&D. Clearly, a larger government size may well be indicating a stronger role of the government in stimulating innovative activity, either through training or provision of funding for research (i.e. an indirect effect). Meanwhile, this may also be an indication of a higher government consumption and tax burden which suppresses productivity (i.e. a direct effect). These findings are relevant to the ongoing debate about the optimal government size since Barro (1991). Another institutional variable that has received much attention from the literature is the strength of private property right protection. Regressions in Table 7 are devoted to examining the impact of private property right protection on productivity. The direct effect of private property rights is found positive but insignificant. Meanwhile, the indirect effect is negative and only significant with the industrial R&D channel. A higher level of private property protection also discourages investment in industrial R&D but has an insignificant effect on how frontier academic research is converted into industrial innovation. These results point out that overly sophisticated and strict private property right regimes may stiffen innovative activity by hindering technological catch-up in countries that have already been innovative (Qian, 2007). This may be because extra protection serves to increase rents accrued to patent holders rather than to reward new innovators (Qiu and Yu, 2010; Sharma et al., 2022). Next, we seek to explore the impact of the conduct of monetary policy on productivity. Results in Table 8 suggest that the conduct of monetary policy has an insignificant direct effect on TFP. By contrast, there is significant evidence that monetary policy positively affects TFP in an indirect way, either via frontier academic research or the industrial R&D channel. This is because a good control of inflation and interest rate is beneficial for long-term research projects (i.e. an indirect effect). Nevertheless, monetary policy is found to negatively affect industrial R&D investment both directly and by reducing the commercialisation of frontier academic research into industrial R&D. Free trade is an important dimension of institutional quality that is widely discussed by economists. Table 9 conducts a test on how free trade impacts productivity in the presence of industrial R&D and frontier academic research. Obtained results indicate that free trade has a positive and significant direct effect, but an insignificant indirect effect, on TFP. In addition, free trade has little impact on the way frontier academic research induces industrial R&D, both in terms of direct effect and indirect effect. These results are in line with the strand of literature characterising the international knowledge diffusion via trade (e.g. Coe et al., 2009). Table 10 concludes the empirical exercise with the use of a measure on limited regulations (i.e. freedom such as business freedom or labour freedom). It can be seen that limited regulations weakly influence TFP either directly or indirectly. However, a reduction in regulations seems to discourage industrial R&D. According to Barbosa and Faria (2011), rigid regulations in the labour market make it harder for firms to flexibly adjust R&D personnel and wages, especially when the wages are sufficiently high. Meanwhile, stringent dismissal laws may encourage firms to provide more training to workers leading to their higher productivity. The offset of these effects will result in an insignificant impact of regulations. The obtained results are in accord with this reasoning. Overall, obtained results confirm that frontier academic knowledge positively affects TFP. While the direct effect is strongly present, there is also evidence suggesting that frontier academic knowledge influences TFP indirectly, specifically via the industrial R&D channel. Institutional quality matters as institutions affect the way academic research is converted into practical innovation in the industry. To a certain extent, institutions also affect the process through which industrial R&D is materialised into technological development. However, different institutional elements affect this process differently. While government size and the conduct of monetary policy positively moderate this process, other dimensions mostly have no significant influence on it.

7. Robustness checks

We perform a number of robustness checks with results being included in the online Appendix. We first collect the research data published by the Times Higher Education (THE) and available from 2011. We use three different research indicators provided by the THE: research scores (RS), citation scores (CS) and research and citation scores (RCS) with the last indicator equal to the sum of the two preceding ones. The correlation matrix in Table A1 indicates a strong correlation between the ARWU publication scores (Pub) and the THE research indicators. With similar regressions as those in Table 4, results reported in Table A2 reveal that while coefficient estimates for some alternative measures of frontier academic research are statistically significant, that for industrial R&D is insignificant across the regressions.We next make use of scores on publications in Nature and Science, the top two journals in science and engineering. This is because it is arguable that science and engineering are the most relevant fields for industrial production. It can be seen from Table A3 that corresponding results are qualitatively the same as those reported in Table 4 that use the publication scores. In our third set of robustness tests, we employ research scores behind the field rankings published by ARWU over 2007–2017.14 To capture the national frontier academic research capability for each country, we create an indicator called STEM, which is equal to the sum of scores on two different fields: Natural Sciences and Mathematics and Engineering/Technology and Computer Sciences. It can be seen from Table A4 that the coefficient estimates for STEM and industrial R&D are mostly similar to those obtained in Table 4. The only difference is that STEM negatively affects SD in column (A4.4). This may be because the time span is not sufficiently long to display any stable long-run relationship between the interested variables.

8. Conclusion

This paper has been concerned with an enquiry into the effect of frontier academic research on technological development, the channel of the impact and the way institutional factors affect these channels. Using a sample of 18 OECD countries over the period of 2003–2017, we found that frontier academic research affects technological change, both directly and indirectly, via the transmission through industrial R&D. Institutions matter as they influence this transmission process as well as the effectiveness of both frontier academic research and industrial innovations on the advancement of TFP. Our obtained results convey several important policy implications. In particular, policymakers should take frontier academic research more seriously in planning their innovation strategies. For instance, it is essential for governments to maintain and grow its support for university-based scientific research, such as ensuring sufficient and reliable funding for academic research, or reducing complex and unnecessary regulations placed on government-funded research projects, to name a few. Furthermore, investment in frontier academic research will work best if it is put in parallel with that in industrial R&D. In order to do so, policymakers should provide more feasible legislation, financial subsidies, policies and other measures to support and strengthen the collaboration between universities and industries. In addition, improving the right type of institutional quality will also enhance innovation and speed up the process that targets at improving the national productivity and achieving better long-term growth. While frontier academic research can significantly induce technological progress, different types of academic research may affect productivity improvement in different ways. For example, it may be interesting to differentiate between research in applied natural science from that in basic natural science and that in humanities and social sciences. Owing to the limited time span of current data, we were not able to perform such investigations and had to reserve this exciting avenue for future research. Another research extension in the future is to examine the effects of institutional designs, those that shape the operation of the market economy such as the electoral system or the regulatory mechanism, besides institutional quality variables that are used in this paper. All these will certainly enrich our future research agenda.

Research in nursing education generates knowledge about student and adult learning and development and provides evidence to guide educational practices. Through research, nurse educators answer important questions about learning and teaching in nursing and test interventions to improve learning outcomes. Rigorous research is essential to build the science of nursing education. Research findings provide evidence that educators need to make informed decisions about learning needs and gaps, curriculum, teaching, assessment, and other areas of education.

Ethical guidelines for conducting research

Research is defined by the Office for Human Research Protections, in 45 CFR §46.102, as the “systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge” (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2018b). Ethical guidelines for conducting research protect the rights of participants and integrity of the research process. The National Institutes of Health (2016) identified seven principles to guide research: the study has value and is important, the research methods are valid and feasible, the selection of subjects is based on the aims, there is a balance of risks and benefits, the study is reviewed by an independent review committee to ensure it is ethically designed and limits conflicts of interest, the participants have sufficient understanding about the study and their role to make an informed decision about their participation, and the processes treat potential subjects and participants with respect. Standards for protecting the rights and safety of participants in research have a foundation in documents such as the World Medical Association’s Declaration of Helsinki (World Medical Association, 2018) and the Belmont Report, published in 1979 (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2016a). In the United States, federal regulations governing biomedical research that involves human subjects are specified in the Federal Policy for the Protection of Human Subjects, known as the Common Rule. These regulations were revised in 2018 (Revised Common Rule) (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2018a). Twenty different federal agencies follow the Revised Common Rule including the Department of Health & Human Services (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2017). To confirm that ethical standards and regulations are met, research studies that involve human subjects must be reviewed by an institutional review board (IRB) (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2018d) or a research ethics committee. The Office for Human Research Protections, in 45 CFR §46.102, defines a human subject as “a living individual about whom an investigator (whether professional or student) conducting research: (i) Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (ii) obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens” (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2018b). Without approval of the IRB or a research ethics committee, the investigator cannot proceed with a study that involves human subjects.

Ethical guidelines for quality improvement

The goal of quality improvement is to improve the quality of care or care processes in a specific setting or organization. Quality improvement is not defined in the Federal Policy for the Protection of Human Subjects (Revised Common Rule) (Bass & Maloy, 2020). However, the Office for Human Research Protections provides a webpage with Frequently Asked Questions to guide QI activities (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2016b). Quality improvement activities “whose purposes are limited to (a) implementing a practice to improve the quality of patient care, and (b) collecting patient or provider data regarding the implementation of the practice for clinical, practical, or administrative purposes” do not meet the definition of research (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2016b). These activities do not need to be reviewed by the IRB.

Educational research and projects qualifying for IRB exemption

In educational research, when data are collected from learners, educators, administrators, and others, these individuals may be considered as research subjects in terms of protecting their rights as participants (Heflin et al., 2016). Those studies, since they involve human subjects, must be reviewed by the IRB (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2018d). However, some categories of research, including education, are eligible for an exemption from the IRB or research ethics committee. Research that is exempt carries no or only minimal risk and meets one of the categories eligible for an exemption. Research conducted in established or commonly accepted educational settings such as schools of nursing on normal educational practices may be exempt under 45 CFR §46.104(d) of the Federal regulations (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2018c). Studies that compare teaching methods, research on variables that influence program outcomes, curriculum and program evaluations, and surveys and tests of students are examples of exempt human subjects research. Only the IRB can decide if a study or project is exempt, not the investigator (Bass & Maloy, 2020; U.S. Department of Health, & Human Services, Office for Human Research Protections, 2019). The determination of exempt status has to be clear and accurate. Because of the investigator’s potential conflict of interest, the Office for Human Research Protections’ recommendation is that investigators should not have “the authority to make an independent determination that human subjects research is exempt” (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2019). However, institutions may have an algorithm, a checklist, or an automated system for certain exemption categories, allowing the investigator to make a decision that the research is exempt based on the criteria established by the institution’s IRB or research ethics committee (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2019).

Guidelines for reporting research

Guidelines for reporting research require that authors of studies involving human subjects state that the IRB or research ethics committee approved the study or determined it was exempt from this approval and why (American Medical Association [AMA], 2020)American Medical Association [AMA], 2020). The AMA guidelines also indicate that investigators “should not make a personal determination about whether a study is excluded or exempt from formal ethical review and approval” and should cite the policies of their academic institution or health care system for the exemption based on the Revised Common Rule (AMA, p. 312). Across all types of studies, IRB or research ethics committee approval should be reported in the manuscript (American Psychological Association [APA], 2020). In a discussion by Thayer et al. (2016) about ethical considerations in medical education research and evaluations, the authors explain that although IRB oversight may not be required, this determination is needed to publish the findings, with most journals requiring this for review of the manuscript. Few studies have examined guidelines of nursing journals and requirements for IRB statements in manuscripts submitted to and published in these journals. A review of author guidelines of 245 nursing journals in the Directory of Nursing Journals, maintained by the International Academy of Nursing Editors (INANE), indicated that most journals had statements about manuscript originality and proper attribution (n = 204, 85.7%), permission to use copyrighted material (n = 204, 85.7%), copyright transfer (n = 194, 81.5%), and conflicts of interest (n = 189, 79.4%) (Oermann et al., 2018). However, that study did not examine requirements for statements about IRB approval.

Purpose

Research has not been conducted on the requirements of nursing journals for educational research, evaluations, and projects to be reviewed and approved by an IRB or research ethics committee nor the types of statements to include in manuscripts. One of the purposes of this study was to describe the requirements of nursing journals for educational studies and quality improvement projects related to education to be reviewed by an IRB or a research ethics committee. The second purpose was to identify the types of statements of approval or exemption to be included in manuscripts submitted to and published in the journal.

Method

Design and sample

The investigators conducted a descriptive study using an electronic survey to collect the data. The sample was comprised of members of the INANE list serve. This list is comprised of editors, associate editors, editorial board members, authors, and publishers of nursing journals throughout the world. Established in 1982, the mission of INANE is to promote best practices in publishing and high standards in the nursing literature. The letter from the investigators invited list serve members to complete the survey if they self-identified as editors of journals that publish educational studies or educational quality improvement projects. Instrument The investigators developed a 10-item survey based on their experience as editors of nursing journals, which is available from the authors. Questions examined the primary purpose of the journal; if the journal published educational studies involving nursing students, faculty, and/ or academic administrators as subjects or educational studies involving nurses as subjects (professional development); if the journal published educational quality improvement projects (related to students, faculty, and/or academic administrators, or to professional development); if the journal required a review by the IRB, an ethics committee, or a designated authority for educational studies and quality improvement projects related to education; and statements of approval or exemption to be included in a manuscript submitted to the journal. Some questions were answered with a Likert type scale, and others were answered through a “check one” or “check all that apply” response. Respondents had an opportunity to add comments for some of the questions. A draft of the survey was distributed to eight experts in the field of publishing for their assessment of clarity and to determine if any topics were missing. A total of eight respondents provided helpful feedback. Those who provided feedback were editorial board members, publishers, editors, and experienced authors. Based on their feedback, the survey was revised.

Procedure

The study was reviewed by two of the authors’ University Institutional Review Boards and granted exempt status. Completion of the survey indicated consent to participate. The survey was prepared using electronic survey software. The link to the survey was distributed to all members of the INANE list, and editors who publish education focused research and quality improvement articles were asked to respond. Education focus was defined as research studies and quality improvement projects related to academic education or to nursing professional development. The survey request was sent a second time to the INANE list after three weeks to help increase the response rate. A total of 64 people responded to the survey.

Data analysis

Data are reported in aggregate to maintain anonymity of respondents. The data were downloaded from the survey into Excel for analysis and were analyzed using descriptive statistics.

Results

Responses representing 64 nursing journals were received. A slight majority of responses (57.81%) represented journals that were focused either primarily (n = 16, 25.00%) or included some articles on academic nursing (n = 21, 32.81%) (Table 1).

Academic nursing

Table 2 provides responses from editors of journals that publish educational research and quality improvement projects concerning academic nursing. The frequency of publishing these articles was evenly distributed across the categories of frequently (n = 22, 34.4%), sometimes (n = 19, 29.7%), and rarely (n = 20, 31.2%). Editors reported that quality improvement projects related to the education of nursing students or administrative practices of the program were published sometimes (n = 22, 34.4%) or rarely (n = 25, 39.1%) in the journal they edited. Regarding human subjects’ review for studies involving students, faculty, or academic administrators, the majority of journals (n = 32, 86.5%) always required a review by an IRB or other ethics committee for research studies, and 17 (45.9%) required the same for educational quality improvement projects (Table 2). One respondent offered that a review by the IRB or other ethics committee was required “for research studies and quality improvement projects that include information directly from the above participants [students, faculty, or academic administrators]. For program evaluations that include only the authors’ observations and experiences, then sometimes no.”

Professional development

While only a few of the journals had a primary purpose related to the education of nurses (professional or staff development) (n = 7, 10.94%), a number of the journals published some articles focused on educating nurses (n = 20, 31.25%). Table 3 indicates that nursing journals frequently (n = 15, 23.44%) or sometimes (n = 27, 42.19%) publish articles on educational studies related to professional development. Quality improvement projects, frequently the focus of clinical settings and Doctor of Nursing Practice projects, are published as reflected in Table 3. Respondents indicated a wide range of responses about publishing such articles – from frequently (n = 10, 15.63%) to never (n = 15, 23.44%). Most commonly, an IRB or a research ethics committee review was always required for educational studies involving nurses (n = 24, 88.9%), and just over half always required a review for quality improvement projects related to the professional development of nurses (n = 14, 51.9%), as shown in Table 3. The open comments revealed that some journals accepted the institutional decision related to not being reviewed because quality improvement was deemed exempt from review. Another view was that quality improvement studies did not require IRB review but needed some kind of authorization. One response indicated such review was expected for any study because learners were viewed as a vulnerable population.

IRB statement in manuscript

Fig. 1 addresses the issue of the expectation for a statement in the manuscript that the study was reviewed or deemed exempt by the IRB, a research ethics committee, or a designated authority prior to the implementation of an educational research study or quality improvement project. The majority of journals require statements for both types of publications, although the proportion requiring a statement for research studies was much higher (89.06% vs 58.33% for quality improvement). Two editors indicated that they accept retrospective IRB approval. For quality improvement, the most common response indicated a statement in the manuscript was required (n = 35, 58.33%), although 14 (23.33%) did not require this and 11 (18.33%) indicated “maybe.” One comment was that such an expectation might not be required if the project related to routine evaluation of development activities. Responses also included acknowledgment that institutional requirements might vary, but an explanation of some type of oversight or determination that no oversight was required was needed. Fig. 2 illustrates the type of IRB statement required by journals. Most required a statement from the author(s) that IRB, research ethics committee, or designated authority approval was obtained (75.00% for research, 61.67% for quality improvement). For research studies, it was more common for editors to request the approval number for the study (18.75%) or the copy of the IRB certificate (18.75%). In addition, one editor responded that the institution granting approval must be named.

Discussion

In many nursing education studies, data are collected from students or other types of learners, educators, administrators, and others involved in the educational process. These individuals may be considered as research subjects in relation to protecting their rights as participants. Educational studies that involve human subjects should be reviewed by an IRB or other type of research ethics committee before implementation of the study. Many educational studies are eligible for an exemption if the study has no or only minimal risk and is conducted in established educational settings on normal educational practices. Research conducted in established or commonly accepted educational settings such as schools of nursing on normal educational practices may be exempt under 45 CFR §46.104(d) of the Federal regulations (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2018c). The determination of exemption, however, should be made by the IRB or research ethics committee, not by the investigator (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2019). Some educational and health care institutions may have a checklist or another system for determining an exemption, allowing the investigator to make this decision (U.S. Department of Health, & Human Services, Office for Human Research Protections, 2019), but without this, the IRB or ethics committee needs to review the proposed study and determine if exempt. This study confirmed that the majority of nursing journals in this sample always required educational research involving nursing students, faculty, or academic administrators, as well as educational studies involving nurses, to be reviewed and approved by the IRB or other research ethics committee. About half of the journals had the same requirement for educational quality improvement projects. For the majority of journals, authors need to include a statement about this review in both types of publications, although the proportion requiring an IRB statement for research studies was much higher (89.06%) than for quality improvement (58.33%). The findings of this study are consistent with guidance from the AMA (2020) and APA (2020) that IRB or research ethics committee approval should be reported in the manuscript. The mixed response regarding requiring IRB, or comparable approval, indicated that quality improvement does not have the same expectations as research where subjects are clearly seen as requiring protection. Quality improvement studies are likely to continue to increase in quantity due to the expectations of groups such as the Magnet Recognition Program™ and Doctor of Nursing Practice programs. As a result, faculty and clinicians might anticipate that more organizations will have a requirement for approval of such studies. Although many of these quality improvement studies might be deemed as exempt, the expectation that they have been reviewed for human subject protection would lend credibility to such studies. One way to assist in this improvement when conducting such studies would be for journals to require a statement acknowledging what the author(s) did to protect subjects and assure quality.

Preparation of nurse educators

Health care educators who conduct educational studies need to understand the Federal regulations and IRB processes (DeMeo et al., 2016; Heflin et al., 2016; Linden et al., 2019). Linden et al. suggest that increased training of education researchers about these would better prepare them for engaging in education studies and projects and may avoid unnecessary delays in IRB review processes. Better preparation of nurse educators for conducting educational research and quality improvement studies of educational programs is also indicated. In nursing, studies have documented the importance of graduate nursing students understanding research ethics and IRB processes. In one study Dols et al. (2017) described an educational strategy in which graduate students conducted mock IRB reviews of submissions with errors and ethical issues as a way of applying principles of research ethics to simulated cases. Other schools of nursing have developed processes to improve students’ understanding about protecting the rights of participants in research and guiding them through the IRB process including quality improvement projects that were not human subjects research (Foote et al., 2015). Clinical organizations most commonly focus on bioethical research rather than professional-focused studies. Studies are needed to examine nurse educators’ understanding of participants in educational research and requirements for conducting these studies and submitting manuscripts to nursing journals.

Implications for nursing education

To facilitate conduct of ethical research of educational processes, schools of nursing might consider developing an infrastructure that supports educational research and quality improvement. At a minimum, this infrastructure could involve facilitating IRB review prior to implementing educational studies and innovations. Such a process may enhance both the conduct and dissemination of work to advance educational practices. For schools with Doctor of Nursing Practice programs, using best practices in the review and conduct of quality improvement projects is essential. Further, use of standardized reporting guidelines such as SQUIRE (2020) for quality improvement and SQUIRE-EDU (2020) for educational quality improvement reporting would facilitate transparency and the potential for replications of educational quality improvement projects (Armstrong, 2019). Educational quality improvement projects involving human subjects need approval through an IRB review or use of an algorithm or a checklist provided by the institution for the investigator to make a decision that the project is exempt.

Limitations

One limitation of this study was the recruitment of respondents using the INANE list serve. Some nursing editors may not participate in INANE, and others may have decided not to respond to the questionnaire. The questionnaire was evaluated by experts prior to its use for content and clarity, but it has no established validity or reliability.

Conclusion

Educational research studies in nursing that involve human subjects need to be reviewed and approved by the IRB or other research ethics committee before implementation. Many of these studies will be eligible for an exemption, but this determination should be made by the IRB or ethics committee, not by the investigator. Some institutions may have a checklist, an algorithm, or an automated system for the investigator to make a decision that the educational research or quality improvement project is exempt; this process should be explained in the manuscript. If there is no such policy and system in an institution, all nursing education studies and quality improvement projects involving human subjects should be submitted to the IRB or comparable committee for review and approval. The majority of nursing journals require this review and a related statement in the manuscript.

This paper responds to the core question: Why has impact assessment research not made more of a difference? Our aim is to try to explain the failure of economic impact assessment as a research management tool in terms of its limitations in providing reflective insights into how research can be improved as a process. Economic assessment leaves unquestioned the institutional context of research, the influences of this context on the research process, and the implications this has for social and economic outcomes (Nelson and Winter, 1982; Biggs 1990; Rajeswari 1995; Hall et al. 2001; Anderson et al., 2002). We suggest that economic assessment could be supplemented with contemporary, innovation systems perspectives, an approach where an appreciation of institutional context and institutional learning is central to analysis and research management procedures. In order to highlight the need for such complementary approaches we present a critique of approaches to economic impact assessment. We focus on economic impact assessment because while the scope of impact assessment and evaluation approaches is very diverse, economic assessment remains the dominant paradigm in the international agricultural research arena, particularly in the research centres of the Consultative Group on International Agricultural Research (CGIAR), (Horton, 1998). The CGIAR centres have been a key mechanism by which the international community has deployed agricultural research for international development. Therefore the way these centres have approached impact assessment and responded to its findings is an important concern. Our purpose is not to try to deny a legitimate role for economic impact assessment. Ex-ante economic impact assessment lends valuable credence to research investment decisions that also hinge on a range of technical, political and policy objectives. Similarly ex-post impact assessment, in an era of growing scepticism over the value of publicly funded agricultural research, provides a politically expedient way of attributing a value to past investments and justifying further support for research. Similarly technology adoption studies are a useful way of tracking and demonstrating the rate and progress of technical change. However, these assessment roles are arguably more politically important than managerially useful. Our central argument is that the measurement of the economic impact has poor diagnostic power. Only when the framework of evaluation is expanded to include an explicit institutional learning agenda will research managers be able to monitor and evolve new ways of addressing goals such as poverty alleviation. We preface our discussion with some definitions, as there is much confusion over the term institutional context and the related concept institutional learning. The main argument then begins with a brief overview of economic assessment approaches. Next, the innovation system concept is introduced as a means of expanding impact assessment procedures to include an analysis of institutional issues. We illustrate this with three case studies where institutional learning has been an important mechanism for achieving better impact. We continue with a discussion of the implications for evaluation and impact assessment using the innovation system concept as a learning framework for research management. We end by observing that such approaches face an up-hill struggle as they contend for legitimacy as a complementary approach with the dominant and deeply entrenched tradition of economic assessment in the international agricultural research community.

2. The institutional and organisational context of research: some definitional points

2.1. Institutions and R&D

The concept of an institution in relation to the R&D process is open to ambiguity in the literature. Different disciplinary conventions define the term in different ways. The ‘‘institutional economists’’ usually adopt the sociological meaning of the term, referring to things that pattern behaviour—routines, norms, shared expectations, and morals (Edquist and Johnson, 1997). The new institutional economists follow a similar convention viewing institutions-as-rules as governance structures that regulate transactions (North, 1990). The sociology of science community also adheres to this strict distinction between institutions and organisations, the latter being viewed as players or actors whose interaction is governed by institutions (rules, norms etc.) (Raina, 2001). The convention in the science and technology policy literature is to use the term institution as an embedded concept (although there is much inconsistency.) This embedded definition refers to the behaviour of physical organisations dealing with research and development (R&D) and economic activity—research centres, universities, private companies, research foundations, farmer’s associations, co-operatives and so forth. This perspective recognises that relationships and interactions between agents have to involve non-price relationships and that while the transaction costs theory of institutions cannot explain the dynamics of such systems, an interactive learning theory of institutions can (Lundvall et al., 2002).

2.2. Institutional context

The institutional context of R&D, therefore, concerns the rules and norms that govern it as a social process of learning. In practice this means the rules and norms governing:

・ how research priorities emerge, are promoted and executed;

・ the role of various actors involved in the production, transfer and use of knowledge;

・ the relationship between the different actors and the factors that affect their relationships;

・ howresearch performance is evaluated and rewarded (incentives), and bywhom;

・ how R&D is held accountable to different interest groups and society as a whole;

・ how knowledge is built up, shared and used; and

・ how organisations reflect and learn.

Other aspects of the institutional context concern the wider institutional environment. For example it affects the way national culture embeds in the norms of individuals and organisations and the way this affects how they operate, interact and relate to each other and how they learn and use knowledge. Therefore there can be different national cultures of science, with norms of acceptable behaviour, review and validation (personal communication Dr S. Biggs). There are also different organisational cultures and traditions in different sectors. For example government agencies (sometimes unfairly) are thought of as top-down bureaucracies, whereas non-government organisations are usually (sometimes incorrectly) presumed to have flatter management structures. These are illustrations of institutional contexts that impact on the way decisions are made, whose voice is heard and the dynamics of relationships with partners—all factors that impinge on the direction and outcome of R&D.

2.3. Institutional learning

Institutional learning, therefore, concerns the process through which new ways of working emerge. It concerns learning how to do things in new ways. It asks the questions: ‘‘What rules and norms have to be changed to do a new task or to do an old one better?’’ ‘‘How has our research approach changed in response to the need to improve the poverty relevance of our work and what else needs to change?’’ and ‘‘What can we learn from activities that did not produce the expected outcomes?’’ A key solution may involve learning how to learn better, a concept that the management and organisational theory literature refers to as double-loop learning (see for example Smith and Stacey, 1997). The learning process is context specific and consequently institutional learning can lead to great diversity of approaches, partnerships and strategies. As our case studies illustrate, institutional learning is an inevitable and intuitive process, a fundamental property of all social systems (ibid.). However where programmes have explicit, systematic learning objectives and procedures, research management strategies can evolve and progress rapidly (Horton, 1999).

3. Impact assessment in the CGIAR: history, conventions and limitations

3.1. Evaluation traditions

The best known impact of the CGIAR system came from its work on developing high-yielding, fertiliser-responsive varieties of rice, wheat and maize. Widespread adoption of these modern varieties in Asia rapidly increased food production. This is commonly referred to as the Green Revolution. The CGIAR has not been complacent about its success. In recent years, with the encouragement of donors, considerable effort has been spent on impact assessment activities see Feldman, 2000. Underlying this has been the feeling that while much has been achieved, the task of reducing poverty still needs more effort and that perhaps approaches that served the CGIAR well in the past may need to be revised in the light of an increasingly complex development scenario. Furthermore, while there is now a considerable history of monitoring the performance of the CGIAR, the fact that it is difficult to detect shifts toward greater impacts has caused a growing sense of unease in the CGIAR and amongst its sponsors. This is increasingly causing some of those involved in impact assessment work to re-examine commonly used evaluation approaches, and in particular the value of these in terms of improving impact. Horton (1998) describes a disciplinarily-diverse typology of research evaluation employed in the CGIAR. In addition to economic evaluation he cites peer review and external review by expert panels as major evaluation types, with bibliometrics, social and environmental impact assessment, and participatory evaluation as minor branches of evaluation. The professional evaluation community brings a richness of disciplinary perspectives and over the years its concerns have broadened from accountability to programme improvement, decision support and institutional learning (Horton, op cit; Horton and Mackay, 1999; Cracknell, 2000). In the CGIAR, however, the economic evaluation tradition is by far the dominant approach, with ex-post assessment dominating ex-ante analysis.

3.2. Audiences for economic impact assessment

The recommendation for international agricultural research has been that impact assessment should be routinely used with two audiences in mind (1) to provide research managers and scientists with information about how technology influences the welfare of agricultural producers and consumers, and to improve targeting of research programmes though adjustments in resource allocation and (2) to provide governments and donors with evidence of the social benefits of investment in particular research programmes and in publicly funded agricultural research generally (Alston et al., 1995; Maredia et al., 2000; TAC, 2000). In the sphere of evaluation there is always going to be a potential dichotomy between accountability and diagnostic objectives (Cracknell, 2000). But since in economic impact assessment of agricultural research the different objectives are rarely made explicit, one can only assume impact research is serving both audiences. For readers unfamiliar with what economic impact assessment means in practice we present a synthesis in, Appendix 1, of the published impact assessments of the International Crop Research Institute for the Semi Arid Tropics (ICRISAT). The synthesis highlights the methods used, the economic impact estimated, and the consequent recommendations made.

3.3. Determinants of the impact discourse in the CGIAR

The reasons why economic impact approaches have come to dominate impact research in international agricultural research are complex. However the publication of Science Under Scarcity: Practices for Agricultural Research Evaluation and Priority Setting (Alston et al., 1995) as well as other ‘‘best practice’’ statements from the international agricultural research community (e.g. TAC, 2000) have helped establish these approaches as the industry standard. Horton (1998), citing key CGIAR documents and declarations, points out how agricultural economics has come to dominate the social sciences in the international centres; the ascendancy of economists to senior positions at a time went the CGIAR was under close scrutiny regarding value for money; and the related emergence of research evaluation as a prestigious area of specialisation within the discipline of agricultural economics. These developments have also greatly influenced thinking in national agricultural research programmes. The discourse on impact assessment has developed as it has because of the specific circumstances that have shaped evaluation conventions within the CGIAR research system. It is informative to recognise where the advocacy for such approaches stems from and thus the types of institutional changes that will be required if complementary approaches to economic impact assessment are to be more widely used in the CGIAR.

3.4. Economic impact assessment and its limitations

Adapting from Maredia et al. (2000), the three main economic impact assessment methods are as follows:

・ Adoption studies/partial impact assessment studies. These do not estimate aggregate benefits, but trace the use of innovations. Adoption studies may also evaluate private benefits in the form of increased farm production and incomes, assess client satisfaction with research results and seek to understand why technology is being used or not. These studies contribute a valuable understanding of the acceptability and performance of technology in given social and physical contexts. Lipton and Longhurst (1989) provide a comprehensive review.

・ The economic surplus approach. This approach was pioneered by Griliches (1958). It estimates the returns on investment, calculating the change in consumer and producer surpluses that result from technological change brought about through research that causes the industry supply function to shift outwards. The estimated economic surpluses, together with research costs are then used to compute the net present value or internal rate of return.

・ The econometric approach. This approach employs a production function, cost function, or total factor productivity analysis to estimate the change in productivity due to investments in research. Analysis is conducted within the framework of a production function that incorporates conventional inputs (land, labour etc.), non-conventional inputs such as education, infrastructure etc.) and the stock of technical knowledge (investments in research and extension). The estimated research coefficient is then used to calculate the value of output attributable to lagged research expenditures and to derive a marginal rate of return to research investment. It is a statistical way of isolating the effect of research on economic output. In international agriculture, Evenson (e.g. Evenson and Pray, 1991) pioneered this approach.

All three economic impact assessment approaches are employed in ex-post analysis, although adoption studies and economic surplus models are the most popular and arguably the most straightforward to use. Ex-ante studies favour the economic surplus model. While these approaches have a useful role there are also limitations and considerable debate exists on methodological issues, particularly for econometric approaches (Maredia et al., 2000). Other acknowledged limitations include: the availability of adequate input and output data on the research process and subsequent technical change (Alston et al., 1995); and the related limitation concerning the difficulty of attributing past, current or future outcomes to particular research investments and assigning a value to these outcomes. Efforts to refine methodologies and resolve the conundrum of wildly differing (but usually high) internal rates of return to research has attracted a great deal of attention from economists (e.g. Alston et al., 1998). The consequences of assessing the impact of ‘‘winners’’ only has also skewed the results of rates of return analysis and possibly undermined the verisimilitude of claims made by econometric studies. Our critique of economic impact assessment does not, however, relate to these methodological limitations. Ours is a more fundamental conceptual problem and concerns the way these approaches exclude the research process and its institutional context from the analysis. As a result, impact assessment makes measurements of research inputs and outputs without measuring and accounting for the content of the research process (Rajeswari, 1995). Economic assessment approaches take the research process as a fixed parameter in the analysis. If this parameter is taken to be stable and working optimally—as economic analysis assumes—then the task of improving the impact of research can be reduced to identifying the most important problem (priorities) on which to focus research, e.g. a particular production pest. Alternatively external conditions that are impeding the productive use of research products can be identified e.g. pricing policies. Identifying both sets of factors has added value. The trouble is that there is no optimal blueprint for the research process as procedures are evolving continuously. Similarly there is a range of institutional issues that shape the outcome of this process in different ways. Taken together this means that economic impact assessment approaches disregard an important set of process and institutional issues that represent a critical arena in which research performance and thus impact can be improved. The next section discusses this omission and ways of planning and evaluating institutional context.

4. The institutional context of R&D and its importance

A recurring feature of the debate on the institutional context of R&D has been the tendency of different arrangements to include or exclude different groups of actors and to determine the role these actors play. For example the traditional convention has been to view scientists as the source of new agricultural knowledge, with this knowledge being delivered to farmers via a separate extension service. The shortcomings of this hierarchy with its narrow set of actors is now the basis of the classic critique used to explain why agricultural R&D has such difficulty in addressing the specific needs of poor rural households and has tended to marginalise their potential contribution to the innovation process (Biggs and Clay, 1981; Rhoads and Booth, 1982; Richards, 1985; Chambers and Ghildyal, 1985.) Over the last two decades policy analysts have consistently urged evaluators and planners to pay more attention to the institutional context and its influence on the content and direction of R&D (Biggs, 1978, 1990, 1995; Biggs and Matseart, 1999). This same theme has informed a growing body of interpretative accounts of agricultural R&D and technology promotion efforts. For example, Anderson, (1991) and Anderson et al. (1991) provide a retrospective analysis of the political economy of international rice research in Asia; Biggs (1982) describes factors shaping international efforts to introduce a new crop into Asia; Clark and Clay (1986) give an account of experimentation with the role of scientist in relation to development actors in a rural development project in India; Greeley (1989) gives an account of the political economy of research and interventions associated with post-harvest grain losses; Rajeswari (1995, 1999) offers a retrospective analysis of the conceptual basis for evaluating agricultural R&D performance in India; Hall and Clark (1995) describe the effect of institutional arrangements on the promotion and diffusion of Rhizobium technology in Thailand; Gass et al. (1997) describe the stakeholder context of rural mechanisation; Alsop and Farrington (1998) offer an account of agricultural R&D as part of multiagency rural development intervention in India; Lewis (2001) describes actor analysis in aquaculture research in Bangladesh; and Bellum et al. (2001) characterize the establishment of new patterns of funding from the private sector at ICRISAT. It is now recognised that agricultural innovations come from multiple sources: research staff; development agencies; farmers; NGOs; private companies; entrepreneurs and artisans (Biggs, 1990). Each set of actors has its own agenda and these agendas may often be divergent and contested. This implies a model of agricultural innovation where interactions between actors are multiple, iterative and evolving, and where the groupings of actors that exist at a given point in time reflect the relative strengths of current political and institutional interest groups. These types of process are well known to many research and development practitioners. However in both national and international agricultural research such concepts have failed to impinge on research evaluation and planning norms. In our view this will continue until a more inclusive analytical framework is accepted as a complementary approach to conventional research management procedures (Hall et al., 2000, 2001).

5. The innovation systems perspective

While perspectives espousing an engagement with institutional context have occupied only a modest amount of attention in the international agricultural research policy community, the perspective has come to dominate the policy debate and practice in other research and economic sectors. It is surprising to find that concepts that are informing international agricultural research policy were superseded a decade ago in this wider science and technology policy arena. Velho (2002) provides a chronology of these developments explaining how innovation systems perspectives have come to the fore. (Ruivo (1994) offers a similar discussion of the changing ‘‘paradigms’’ of science policy). The contemporary debate from this parallel policy literature now takes it as given that the linear model of innovation is of little value in evaluating and planning R&D. There has been a shift in the role of policy from examining the determinants and consequence of research, to an innovation role where emphasis is on strengthening networks of users and producers of knowledge (Velho, op. cit.). Underpinning this shift of perspective over the last two decades has been a deepening understanding of the nature of innovation as a process and the accompanying realisation that neo-classical economics alone cannot explain the dynamics of economic systems. There are many avenues of thought that have supported this emergent view. One is the empirical work of Freeman (1987) and others on the institutional arrangements associated with innovation performance. Another is the evolutionary economics perspective of for example, Nelson and Winter (1982). This argues that human behaviour is not characterised by processes of maximisation but instead, over time, the notable feature is learning and change. Similarly the complex systems ideas that explain human behaviour (including R&D and economic performance) in terms of the boundary state, instability and the consequences for organisational and institutional learning has highlighted the value of considering development and change in systems terms (Smith and Stacey, 1997, Clark et al., 1995). Another way of thinking about innovation is that proposed by Gibbons et al (1994) in their discussion of two modes of knowledge production. In mode one, knowledge is generated, often with government assistance, by a research community accountable to its disciplinary peers. Gibbons’ posits that institutional changes in western societies (where the market has started to eclipse the state as the primary decision-maker) have forced science to become more socially embedded and less hierarchical, thus defining the mode two type of knowledge production. As societies and economic systems become ever more complex, the mode one type of production of knowledge will become less able to respond to rapidly changing user contexts. Only by assuming the features of mode two can systems cope with complexity and rapid change. The innovation system concept serves to draw some of these ideas together. The ideas of a ‘‘national system of innovation’’ (Freeman, 1987; Lundvall, 1992) and related frameworks (Edquist (1997) and Andersen et al. (2002)) have had considerable influence in the policy analyses of institutional systems that underpin innovation. At its simplest the concept recognises that innovations emerge from systems of actors. These systems are embedded in an institutional context that determines how individual actors behave and how they interact with other elements of the system. Lundvall (1992) identifies learning and the role of institutions as the critical components of such systems. He considers learning to be an interactive and thus socially-embedded process, which cannot be understood without reference to its institutional and cultural contexts. Successful systems are characterised by:

・ Continuous evolutionary cycles of learning and innovation;

・ combinations of technical and institutional innovations;

・ interaction of diverse research and non-research actors;

・ shifting roles for information producers, information users and transfers of knowledge dependent on a need basis

・ an institutional context that supports interactions, learning and knowledge flows between actors.

The application of this concept of a national system of innovation framework in the agricultural research sector is gaining ground (Hall et al., 1998, 2000, 2001, 2002a,b; Rasheed Sulaiman and Hall, 2002; Mehra, 2000; Clark, 2002; Clarke et al., 2001; Ekboir and Parellada, 2001). At the heart of this framework is the contention that R&D is always embedded in social, political and institutional contexts and that unless the influence of this environment is accounted for by decision makers, the evaluation and planning of R&D will be incomplete. What does this mean for the evaluation and planning process? Some of the principles that are required to relate R&D to institutional context include:

(i) An inventory of innovation actors. The framework provides a starting point for identifying the full range of actors relevant to a particular innovation system. While many of the normal public-sector actors are present in the conventional policy schema, closer investigation reveals a wider range of individuals and organisations from other sectors.

(ii) System competency. Once a full inventory of actors has been established it is then possible to examine the extent to which relationships exist among actors. The existence of relationships will depend on the policy context and the wider institutional environment. For example, strong public-private partnerships may have emerged through a liberal policy towards germplasm access. Alternatively, weak linkages may be a result of restrictive personnel polices for public sector scientists that prevent them undertaking contract research for the private sector. Hence analysis has the effect of directing the focus of evaluation and planning on linkages that need to be developed and on potential policy changes.

(iii) Actor roles. Part of the relationship analysis concerns the importance of multiple roles played by some actors and the different types of relationship these roles imply. For example, an agricultural university may be both a source of information on regional variety trials, as well as a recipient of improved breeding lines from a crop improvement centre. Both types of role are important for an effective innovation system and the evaluation and planning process needs to understand their separate but linked existences. Actors with important roles that are excluded from existing arrangements need to be recognised. Technology users and product consumers from poor communities are examples.

(iv) Cultural context. The types of relationship that develop in a particular innovation system reflect both the national context as well as different organisational cultures. For example the national context may have for historical reasons a strongly paternalistic public sector culture with a mistrust of private sector enterprise. Or the public sector may have a strongly hierarchical culture, whereas the NGO sector may have a more decentralised, participatory culture. Partnerships between public agencies and NGOs will not necessarily lead to more participatory approaches, because of the organisational culture of the former. The evaluation and planning process needs to account for these contextual features.

(v) Relationship dynamics. The importance of the nature and dynamics of relationships between the entire range of actors, from the innovation systems point of view, is that their analysis reveals that such relationships are often strongly asymmetrical, preventing interactive learning. For example, partnerships between international and national agencies are often skewed by more favourable access to resources on the part of the former, by historical patterns of interaction, and by professional and cultural norms that value ‘‘outsiders’’ at the expense of ‘‘locals’’. Local political processes, interest groups, ethnic communities, and social hierarchies will all contribute to the political economy of the innovation process. The evaluation and planning process will benefit from an awareness of these dynamics.

(vi) Reflection and institutional learning. The innovation systems framework regards reflection on process and institutional learning as key elements for success. For example, systems in which there is clearly a gulf between policy rhetoric and research practice have a weakness with regard to institutional learning. Other indicators of weak institutional learning may be a reluctance to admit mistakes and confront failure and its causes, or even a reluctance to revisit key assumptions about roles or ways of working. In contrast, organisations in which senior management encourage and reward reflection and learning and where self-evaluation is undertaken regularly, demonstrate a tendency to possess a higher capacity for continuous institutional learning and innovation. The evaluation and planning process could benefit from recognising the importance of a learning culture within public-sector research organisations and their partners.

The innovation systems framework is a learning framework. This characteristic makes it critically different from conventional frameworks (such as the project cycle, of which economic impact assessment is a part) which are problem-solving frameworks. In the next section we present three case studies that demonstrate the role of institutional learning, and the way it affects the innovation systems involved.

6. Case studies of institutional learning

6.1. Case study 1. Tacit lesson learningby scientists

Based on Hall (in press), this first case study explores how scientists working on a long-term sorghum and millet improvement programme (SMIP) learned in a tacit way the value of working in alliances with new partners. This tacit learning was then used as a response to the increasingly impact-oriented agenda of international agricultural research. SMIP is a 20 years initiative supported by United States Agency for International Development (USAID) and implemented by ICRISAT on behalf of the Southern Africa Development Community. It started in 1983 and was implemented in four five-year phases, the fourth running from 1998 to 2003. The first two phases concentrated on developing research infrastructure and human resources in the national agricultural research organisations (NAROs) of the region. This involved the establishment of breeding programs, including research infrastructure and the sponsorship of doctoral and vocational training for scientists. It was done with a view to building capacity to produce a stream of technologies, mainly improved varieties. Indeed, during these first two phases considerable technology development work took place, with 15 varieties being released. The third phase (SMIP III) 1993–1998, while continuing capacity building and technology development activities, started to shift focus towards technology transfer. This change related to developments in research methodology, particularly farming systems and participatory approaches, and the way these developments were starting to influence the thinking and agenda of SMIP. An equally important influence was the wider political economy of international agricultural research at that time; in particular the growing disillusionment among donors and an increased scrutiny of the impacts of research efforts that they were supporting. During phase III, SMIP began to engage in partnerships with actors other than national agricultural research organisations (NAROs). This was in response to the need to have more direct contact with farm communities and the perceived value of working with non-government organisations (NGOs) as a means of achieving this. Analysis of constraints to adoption of technology had highlighted weaknesses in variety release and dissemination systems. It also became increasingly apparent that to achieve wider improvements in seed systems (as well as in other spheres), SMIP and NARO scientists would have to link with a range of other partners—the private sector, including NGOs, and community based organisations (CBOs). SMIP phase IV was seen by the donor, USAID, as a way of capitalising on earlier investments in capacity building, research and technology development. This technology transfer theme meant that SMIP would need to continue to broaden its focus beyond conventional scientific activities and the generation of new technology, instead adopting a stronger developmental focus. Pursuing these developmental goals through a broader range of partnerships became an explicit objective. The developmental focus and the partnership approach were reinforced by the USAID-style project structure and its monitoring procedure. This entailed the identification of a number of intermediate results. Not only were these prioritised by a group of regional stakeholders, but also the quantitative indicators for the achievement of these intermediate results were defined, with annual targets set to monitor performance. The SMIP scientists leading the programme component under each intermediate result became directly accountable for achieving these targets. These included: area sown to new varieties; tonnes of sorghum and millet entering commercial markets in key locations; quantity of seed produced of new varieties. This pattern of accountability was a significant new feature of SMIP IV. The four SMIP scientists quickly realised, based on past experience, that if they were to achieve these targets an entrepreneurial approach to partnership would be essential. An institutional analysis undertaken after two years of phase IV (Hall, in press) found that the SMIP scientists had entered into a broad range of partnerships with both NGO and commercial sectors, as well as with their conventional NARO partners. Drawing together clusters of partners around specific themes or tasks had been used as a way of achieving targets. Hall referred to these as task networks and noted:

・ SMIP scientists played multiple and different roles in these task networks— sometimes as facilitator, sometimes as a source of information, sometimes as the recipient of information,

・ constituent actors were specific to a task theme (due to resources, interests and agendas), as well as to a particular location and institutional context (who were available and how their interaction was governed);

・ there was evidence of the task networks as a mechanism for priority setting for further research, but this was limited and had not been exploited;

・ SMIP’s task networks appeared to represent new forms of collective capacity and important lessons could be drawn from this and shared with scientist and research managers at both ICRISAT and in the Southern Africa region;

・ opportunities for learning institutional lessons and promoting them more widely were restricted by the overall problem-solving, output-orientated framework of the project design (attributable largely to the donor), and the limited formal opportunities this presented for systematic learning.

This case raises a number of points about how new research procedures evolve. The first concerns the complementarity between formal adoption studies that suggested the need to work with new partners in seed systems and the intuitive, tacit learning process through which scientists built up experiences of working in a new partnership mode. The second concerns how the skills and lessons that scientists had built up surrounding the diversification of their partnership base were used to respond to a major institutional change in terms of accountability. This led to a major research innovation for ICRISAT. The third point, however, is that because the overall research management framework focused on monitoring progress in conventional impact terms, this innovation remained largely unrecognised (except among the scientists involved) and did not bring about wider changes in research practice in ICRISAT or its NARO partners. An important implication is that conventional monitoring systems like the one in place in this case while sufficient for accountability purposes to the donor, fail to capture, synthesise and report important institutional innovations that would appear to correlate strongly with achieving research impact.

6.1.1. Implications for impact assessment procedures This case highlights the fact that whether or not projects or programmes are specifically designed to generate institutional lessons, learning is an integral part of any task and the social process of research will inevitably generate such institutional innovations. It also points to the fact that the promotion and diffusion of these innovations requires explicit incentives to be in place. Requirements include incentives from those sponsoring research as well those from within the organisations conducting research. To do this will require explicit institutional changes in the areas of research funding, planning and execution. Changes are required in the conventions of research management and research evaluation. These changes must legitimise and encourage the (i) discussion, (ii) the documentation and (iii) the promotion of institutional innovations and grant them the same level of attention and importance as technological achievements enjoy.

6.2. Case study 2. Learningby confrontingorg anisational culture

The second case explores how ICRISAT dealt with a major decision about its evolving relationship with the Indian private-sector seed industry and the how this has opened the way for a variety of relationships with partners (Bellum et al., 2001). The origins of these developments stems from the liberalisation of the Indian seed sector in 1988 and the consequent emergence of private seed companies in what had previously been a state-run domain. ICRISAT supported fledgling sorghum and millet seed companies and provided breeding material. Support was often through informal networks whereby scientists trained at the Institute found employment opportunities in the new rapidly-expanding private seed sector. During the 1990s these companies began to develop their own R&D capacity, but continued to value the improved breeding lines from ICRISAT. However ICRISAT’s breeding strategy altered in the mid to late 1990s, switching its emphasis to traits suitable for African production and consumption contexts. These were different from the traits Indian seed companies had identified as the priorities of Indian farmers and consumers. As a result the seed industry had to seek new ways to develop hybrids suitable for Indian production conditions and consumption preferences. At the same time ICRISAT began to recognise that the private sector was likely to become a major mechanism for delivering improved breeding material to farmers and that in all likelihood this would prove a more effective mechanism than the public-sector seed system. A number of ICRISAT breeders realised that the way forward was to enter into a new form of relationship with the private sector. This meant a shift from viewing seed companies as passive recipients of breeding material, to viewing them as active research partners and as a source of research funds. There was some caution on the part of the seed industry. The key breakthrough, however, was the suggestion made by the then President of the All India Seed Association to orchestrate funding through a consortium of private seed companies. This helped reduce the cost for individual companies, and broad-based membership avoided the risk of ‘‘free riders’’. The Intellectual Property Rights policy of ICRISAT is such that it does not take ownership of the material it develops. It transfers material to others under a formal transfer agreement which requires that recipients also forgo ownership (ICRISAT, 2001). Such a requirement makes exclusive arrangements with a private company difficult. The consortium approach was one way of addressing this problem. Another was by interpreting the material rights agreement in such a way that if ICRISAT material was further developed by the private sector companies by combining it with their own breeding lines, they could claim ownership of the new varieties produced. Having convinced the private sector that funding ICRISAT was the way forward, ICRISAT breeders then faced the task of convincing ICRISAT management that partnership with the private sector was appropriate. These negotiations started at a time when the mandate of the Institute was still interpreted in a highly circumscribed fashion, based on a rigid notion of the nature of its international public good role. No previous agreement had been entered into whereby the private sector funded research at ICRISAT. There was a perception within the Institute—albeit never explicitly articulated—that its public good role could only be maintained through purely public funding and execution of research. In part, this perception was informed by the political realities of ICRISAT’s (often highly sensitive) relationship with the Indian public-sector agriculture research organisation through which it is mandated to work. The result was that a final decision to approve the consortium proposal could not be made at the Institute level. Responsibility for the decision was passed to ICRISAT’s Governing Broad, who in turn passed it to the Indian Council for Agricultural Research. The Council approved the proposal and passed it back to the ICRISAT Director General at a time when a previous Director General from the 1970s and 1980s had returned on an interim basis following the sudden departure of the incumbent. The proposal was rejected by the Interim Director General apparently because of the relatively small sums of money involved and administrative concerns about managing small grants. Finally when a new ICRISAT Director General was appointed, he approved the proposal stating that its importance resided in the new partnerships involved rather than the financial considerations alone. This process had taken almost 2 years. The consortium approach has subsequently evolved with sufficient success to attract additional private-sector seed companies to join the consortium and extend funding through this mechanism. The impact this has had on the organisational culture of the Institute has been widely felt. Joint initiatives with the private sector are now viewed as relevant to the broader developmental mandate of ICRISAT. It has lead to the development of a ‘‘Technology Innovation Centre’’ that acts as a clustering device for a range of special projects that involve new types of partnerships and relationships. For example sorghum and millet breeders are now pursuing relationships with private animal-feed industries as a means of developing new market opportunities for crops for which traditional food uses are declining. Other initiatives include an incubator for Indian biotechnology companies and linkages to a newly established science park in the nearby state capital. The special project status of the Technology Innovation Centre allows new institutional arrangements to be tested; for example cost recovery and profit sharing, joint development of research priorities with partners.

6.2.1. Implications for impact assessment procedures

This case highlights the way a new type of interaction with partners has allowed the Institute as a whole to learn new ways of operating and the way this has resulted in capacity development. Specifically it has changed the capacity of the organisation to work in a more interactive, less hierarchical way with its partners. In addition, new capacity exists as a result of the novel combinations of resources and expertise of ICRISAT and its new partners. In this case these new capacities will lead to improved impacts through the better delivery of research products to farmers. Impact assessment should be playing a role in strengthening this capacity development process, helping formalise the learning process and promoting the lessons learnt, as well as highlighting the value of these capacity outcomes in internal monitoring procedures. Another implication of this case is that impact assessment procedures must be undertaken jointly with partners who form part of this new capacity.

6.3. Case study 3. Learningas a way of dealingwith the institutional context of research

The third case study discusses how the crop post-harvest programme of the Department for International Development (DFID), the UK government’s international development assistance agency, has gradually recognised the need to pay more attention to the institutional context of the research it was sponsoring and how it responded with an approach that is attempting to embed institutional learning in conventional technology-development projects (Hall and Rasheed Sulaiman, 2002). The programme is one of 10 natural-resources research programmes. These were originally established by DFID in 1995 as a way of exploiting the UK science base in support of international development. The programmes were conceived in the problem-solving framework of the project cycle with the ‘‘logical framework’’ used as the key programme and project planning and evaluation tool. This was supplemented by monitoring indicators used to judge progress along a notional output pathway. The translation of technical outputs into poverty/developmental impacts was dealt with as a logframe assumption about the existence of ‘‘target institutions’’ (meaning, in this instance organisations) and functioning ‘‘up-take pathways’’. As projects progressed the Crop Post Harvest Programme started to recognise that process and institutional issues were having serious consequences for the success of its research initiatives. For example, in a series of projects commissioned in India to provide technical backstopping to parts of the export horticulture sector, it became apparent that the real problem was one of mobilising the different parts of the public-sector research system to act in a concerted fashion. Collaboration was particularly important for export development because of the need to deal with quality management issues in an integrated production and post-harvest supply chain. In addition, the broad range of stakeholders in the supply chain, including farmers, whose agendas and circumstances provided the context for developing these solutions, made it difficult for the research organisations to respond effectively, given their prevailing way of working with stakeholders.

At this point the programme management team took decided to try to gain a systematic understanding about the way this institutional context was affecting its research. The learning process built up slowly. First there was a pilot project that continued its focus on export horticulture, but which included a simultaneous technical and institutional analysis. This highlighted the need to identify a conceptual framework to help understand the wider contextual issues that were affecting the research process. It was at this point that the programme started to explore the innovation systems framework. The exploration began with a policy project in India to examine how the innovation systems idea could be used in the evaluation and planning of R&D. This project was undertaken with a view to drawing both project and programme management level lessons. It was contingent on the wider programme portfolio of projects in India which in effect acted as case studies. This approach allowed the programme in South Asia to experiment with the innovation system idea, while allowing conventional projects to proceed. It became apparent that the arrangement was not ideal. Notably the institutional lessons that the policy project was gathering from the rest of the portfolio could not be used to redirect these projects as the portfolio was not structured in a truly action-research framework. It soon became apparent that the individual technical projects needed to concentrate on generating their own process and institutional lessons, for project management purposes as well to gain insights of value to the wider programme. However, it was difficult for projects that had been commissioned to deliver a narrower set of outputs, to accommodate this expanded role. Nevertheless, the programme was able to identify and document a series of research management lessons. These included the following:

・ There is a need to build stronger and more consultative linkages between public sector science and other actors in the innovation system;

・ successful projects were those that focused specifically on establishing a coalition of local actors around a particular problem area;

・ these actors included scientists, but not exclusively, and not necessarily as the lead actor. Moreover, roles may evolve over time;

・ the selection of the most appropriate actor grouping was very often an empirical issue that could not realistically be resolved at the outset of a project;

・ There was a tendency, reinforced by the output-oriented, problem-solving framework of the conventional project cycle, to under-report process lessons associated with technological success (or failure.) These lessons were often complementary to new technical knowledge;

・ the relative degree of poverty focus was related to the agendas of different project partners and the dynamics that determined how these agendas were promoted in the wider arena of the project;

・ Needs assessment and participatory approaches were much less important in ensuring a poverty focus than the agendas of the stakeholder involved in projects.

The programme consolidated these types of lesson through a programme-commissioned formative review (Biggs and Underwood, 2001). The review was principally concerned with providing a basis to argue for changes in the programme logframe. Specifically, there was good reason to challenge the need to monitor direct poverty impacts at the project and programme level (even though in the long-term the programme and DFID would be accountable for these outcomes). A more pragmatic approach appeared to be to track behavioural (and therefore institutional) changes that the programme was stimulating among project partners as milestones towards reducing poverty. The key leading indicator thus became the extent to which a systems capacity to innovate in a pro-poor fashion was being developed. The review recommended that to contribute to the development of this capacity the programme needed to:

・ Shift to an innovation systems approach because the emphasis had to move from a problem-solving framework to a learning framework;

・ shift to action research protocols rather than the project cycle management tools;

・ develop projects that involve groupings of local partners (coalitions), where identifying partners becomes part of the research task;

・ use stakeholder analysis to make agendas transparent;

・ monitor partner and stakeholder roles and interests to maintain a poverty focus.

These broad principles have informed programme strategic plans for 2002–2005. As the programme works through some of the wider implications of this shift, it and its project partners will have to continue to use institutional learning as a core research management tool.

6.3.1. Implications for impact assessment procedures

In this third case a fundamental shift in the research-management approach of a donor research programme took place in order to more effectively deal with the institutional context of the research it was commissioning. The programme purpose is no longer solely concerned with narrowly defined impacts on poor people, but rather with the creation of ‘‘post-harvest innovation systems that respond to the needs of the poor more effectively’’. This is important because it shifts the performance of the innovation system out of the assumption column of the logframe. Thus bringing about changes in this system becomes the central endeavour of the programme. There are two implications for impact assessment procedures. Firstly, evaluation for external accountability purposes must include systems and capacity changes along with judgements of conventional impacts on the poor. Secondly, because the programme has adopted an action research approach, monitoring and internal evaluation become critical learning tools for the projects’ own management purposes. Since the projects are operating through coalitions of partners monitoring and evaluation need to be undertaken jointly by the partners as it is their collective capacity and learning which is at stake. The programme’s experiences of implementing this approach highlights the need to build up process monitoring skills among both scientific and non-scientific partners. The wider implication is that if impact assessment is to be reoriented towards a greater emphasis on learning, all actors associated with the innovation process will need to develop new skills to allow them to learn more effectively. This ability to learn will be an important indicator of the emergence of new and more effective innovation system capacities.

7. Lessons and remaining challenges

These three short case studies tend to confirm critical features of the innovation systems concept discussed earlier. Four principal features emerge. Firstly, research is an inherently social process where learningand institutional innovations are part and parcel of technology development and promotion. The first case study illustrated how scientists were the source of this learning, derived from their own experience of trying to get research products to technology users. The second illustrates how institutional lessons were learnt by scientists and administrators contesting critical aspects of organisational culture and how this has led to the development of new capacity to work interactively with a range of partners. The third case illustrates an attempt to embed these institutional learning processes within the boundaries of the research project and legitimise institutional innovations as project outputs. Second, research approaches and outcomes are intimately related to institutional contexts. In the first case study, institutional changes relating to the accountability of scientists, in combination with the technology transfer agenda of the donor, were very clearly related to the strategy used to implement the project. However the institutional context tended to restrict the formal learning, promotion and diffusion of institutional innovations that emerged. In the second case, the consortium funding arrangement would have been rejected if the prevailing norms within ICRISAT had not been contested and changed. Once the institutional context was altered a series of new research partnership possibilities started to become feasible. In the third case the prevailing institutional context was such that rural households had very little influence on the research process despite being, at least in the rhetoric, the principal stakeholder. The response of the donor research programme has been to try to formalise changes in this institutional context in ways that will make innovation pro-poor. Thirdly, the institutional context of research is principally played out in the combinations of actors involved in research and the patterns of relationships between these actors. In the first case study the critical innovation was the shift to a broader-based partnership approach (although this was a response to the wider institutional context imposed by donors). This move capitalised on the developmental and entrepreneurial agendas of NGOs and the private sector. The second case concerned the emergence of the need for a new type of relationship with the private sector. In this arrangement, for the first time a ICRISAT, the research priorities were set by the private sector and the research executed by the public sector. In the third case, the selection of appropriate project partners and exploration of their interests and the nature of their relationships with others became a central mechanism for developing pro-poor innovation capacity. Fourthly, a feature of all three case studies is that capacity to innovate is the combined function of the actors involved, the skills they bringto partnerships and the institutional contexts that shape the interrelationships. The first two case studies in particular describe the emergence of new capacities less concerned with ICRISAT alone than with the capacity of a grouping of partners and how the grouping evolves and changes ways of operating. The third case is an example of experimentation with ways of promoting the development of innovation systems capacities that are pro-poor. One aspect of these new innovation capacities is the ability of organizations to learn. The first two cases illustrate the way learning was intuitive and informal and the way, as a result, lessons spread slowly with changes taking place over an extended period. The third case suggests that if learning is to be made part of the formal mandate of research projects with a (view to developing new innovation capacities), resources will have to be devoted to develop the learning skills of all project partners. An implication of this is that learning skills are likely to be an important indicator of capacity of an innovation system. The case studies therefore provide reason to believe that the concept of an innovation system offers a framework for thinking about research and impact as part of a wider learning process. This has implications for the way evaluation and impact assessment is conducted by the international agricultural research community. To preface these implications we revisit some of the fundamental philosophical underpinnings of the nature of knowledge and the way that it is produced and used. The principal point raised by the innovation systems perspective is that the nature and value of knowledge cannot be viewed as independent from the processes that produce and use it. The corollary is that to judge the value or impact of new knowledge requires an understanding of knowledge production and use contexts. It is this contextual information, typically institutional in nature, that determines outcomes and impacts. Once this position is accepted then evaluation and impact assessment assumes an importance greater than the resource allocation role of economic assessment. It becomes the principal mechanism for strengthening social learning processes that allow organisations to accomplish new tasks and mandates - such as achieving impact or becoming more poverty-relevant. Our case studies have highlighted some of the practical changes that this philosophical shift towards institutional learning and change entails in international agricultural research organisations. These include:

・ moving the focus of impact and evaluation from examining changes in technology user groups to including changes in the way the research community operates as well as its interaction with other organisations and institutional (including political) contexts;

・ introducing institutional changes that provide incentives to formalise learning as part of the practice of research organisations. This requires changes among donors and senior managers of research organisations and probably within professional bodies relevant to the international agricultural research community;

・ recognising capacity development as an important outcome and purpose of research;

・ accepting the need to explore behavioural changes in innovation systems as a way of monitoring progress and learning, as well as a way of promoting critical institutional lessons to wider audiences in the R&D community;

・ recognising the systems nature of capacity development so that evaluation becomes a task that needs to be done collectively with partners as well as at the individual organisational level;

・ accepting the need to embed evaluation as learning in the day-to-day procedures of research staff and administrators and acknowledging the skill and resource implications of this. This implies the need for greater numbers of social scientists in international agricultural research organisations, but with a hands-on role of facilitating learning in addition to disciplinary research contributions. It also implies the need to build learning skills among all partners and to allocate time within the research process for collective learning and reflection.

We do not present the innovation systems framework as a panacea for improving the performance of agricultural research. Our aim is to draw to the attention of planners, evaluators and research managers the need for (and the possibility of) thinking about agricultural research in a more holistic and evolutionary fashion. We see three major challenges for the innovation systems framework. Firstly we have yet to see how institutional learning has led to new stakeholder-driven ways of setting technical research priorities. Our first case showed limited evidence that this can happen when new problems of an applied nature emerge in task networks. Our second case showed that new partners can bring research priorities with them. In the third case future research priorities remain an empirical question as do the processes to negotiate them in project coalitions. Further exploration of this aspect is clearly required, and institutional experimentation to explore this issue specifically would be useful. Secondly, since the innovation systems framework recognises that priorities and agendas are negotiated and contested, greater analytical attention will need to be given to actor interaction and dynamics. These issues have not been explored in great detail in the case studies. To do so systematically will certainly require a broader range of social science tools than is currently employed in conventional research planning and evaluation approaches. It will also require the deployment of considerable international agricultural research resources for skill development and the expansion of the number of social scientists with appropriate analytical perspectives. Fortunately many of the necessary analytical tools already exist. For example practical ways of dealing with the need to address agency roles of different actors has a long history in the action-research tradition (Whyte, 1991). More recent approaches include stakeholder analysis (Grimble and Wellard, 1997). In the context of agricultural R&D similar principles have underpinned a number of recent approaches. For example, the contending coalitions framework (Biggs and Smith, 1998) has been proposed as a way of complementing R&D planning cognisant of the political economy in which multiple (actor) sources of innovation sit. The actor linkage matrix (Biggs and Matsreat, 1999) provides a practical tool to analyse the relationships that surround capacity-building efforts in natural resources R&D. Horton et al. (2000) discuss an action research approach using an organisational assessment framework to understand capacity building in planning, monitoring and evaluation. Horton (1998) points out that the evaluation community has a rich array of tools and disciplinary perspectives. We note with some humility that many of the implications that the innovation systems framework has for impact assessment are similar to the combined learning of decades of evaluation practice (see Horton, 2002). The third challenge therefore does not concern methodological developments to support the adoption of an innovation system framework. It concerns, rather, institutional developments in the international agricultural research community and the need to contest economic analysis as the impact assessment method of choice. Without the legitimization of the innovations systems framework and related learning-based evaluation approaches, agricultural science will remain stuck in repetitive cycles of project implementation and output evaluation. Bereft of learning, it will fail to find better ways to fulfil the social and economic purpose that its significant potential promises.

Teaching and research are central to the delivery of higher education. Indeed, for many observers, it is the relationship between teaching and research which is fundamental in defining the distinctive nature of the university as an institution. Yet, at the same time, this relationship––the so-called teaching: research nexus––is commonly misunderstood and/or is based on unconvincing or conflicting evidence, both theoretical and empirical.

Many papers have now been written using both quantitative and qualitative methods to explore the nature of the teaching:research nexus; many views have been expressed, some suggesting a strong, symbiotic link and others arguing that there is no real relationship in practice between these two key drivers of higher education activity (see, for example, Breen & Lindsay, 1999; Brew & Boud, 1995a, b; Brew, 1999; Brew, 2001; Clark, 1997; Coate, Barnett, & Williams, 2001; Colbeck, 1998; Elton, 2001; Gibbs, 1995; Hattie & Marsh, 1996; HEFCE, 2000; Jenkins, Blackman, Lindsay, & Paton-Salzberg, 1998; Jenson, 1988; Kyvik & Smeby, 1986; Marsh & Hattie, 2002; Neumann, 1992, 1994, 1996; Ramsden & Moses, 1992; Robertson & Bond, 2001; Rowland, 1996; Smeby, 1998). These authors reflect many different research approaches and provide a wide range of different conclusions regarding the nature of the teaching:research nexus. Particular attention should be drawn to the work of Neumann in applying the term ‘‘nexus’’ to the relationship between teaching and research, and to the papers produced for two international colloquia on Research and Teaching held in the UK in 2000 and 2004 (Southampton Solent University, 2004). This paper aims to offer a different perspective, commonly overlooked in the debate to date, that of institutional management. It aims to examine the factors which influence management of the teaching:research nexus by higher education managers and by academic staff, and, thereby, to derive a new conceptual model which will enhance understanding of how the relationship between teaching and research may be organised and influenced in practice. Whilst any model based on four institutional case studies may be of limited relevance in other institutions and other contextual settings, it is suggested that some of the issues raised will contribute to a wider debate on the relationship between teaching and research in higher education. Few researchers have considered the management implications of the teaching: research nexus. Neumann (1993) considered the role of several academic administrators. Another important contribution is offered by Jenkins, Breen, and Lindsay (2003). However, their work is essentially practical in nature, using a range of case studies to indicate possible policy developments. Papers presented at the two international colloquia in Southampton offer many important insights, especially in terms of national policy, but do not offer any conceptual view of institutional management of the relationship between teaching and research. This paper takes an alternative approach, using grounded theory to consider how different factors can influence how research is managed, both by professional institutional managers and by academic staff (‘‘the managed’’). It aims to offer a new, more theoretical and conceptual basis for the activities undertaken by universities in shaping the interaction between teaching and research. It also seeks to move debate in a new direction, towards institutional case studies and the establishment of management concepts to go alongside educational practice.

Research project

The paper is based on the results of a comparative study looking at policy, perceptions and management relating to the teaching:research nexus in four universities, two in England and two in Sweden. Given that a key aspect of the project was to examine the underlying views and beliefs influencing higher education management, a qualitative approach was adopted. The four universities were selected on the basis of their publicly expressed commitment to the interaction of teaching and research and their breadth of academic disciplines; all four institutions draw a high proportion of their funding from public sources. The four universities vary in size from about 15,000 students (University A) to about 25,000 students (University C) (2004 figures). A series of interviews and focus group discussions was undertaken with academic staff, Heads of Department, Deans and senior officers, both academic and administrative, including Vice-Chancellors and Rectors:

Interviews were also undertaken with senior members of the Higher Education Funding Council for England (HEFCE) and the Hogskoleverket and the Higher Education Audit Agency of Government in Sweden. In addition, a detailed study was undertaken of diverse plans and reports published by each university and by Government in both England and Sweden. Having initially sought permission from each university to undertake the study, each institution was invited to nominate a senior manager to assist with practicalities, such as the arrangement of interviews and circulation of information. This person circulated management and academic staff inviting those interested to participate in the study. Once these volunteers had been identified, a final list of interviewees was devised; academic staff were drawn from a range of different subject backgrounds in each institution, in order to allow for variations in perceptions and practice by disciplines, and from different levels of seniority. Thus, the number of participants varied between different institutions. A short paper outlining the issues to be discussed was sent to all participants 2 weeks before the meetings. This paper formed the basis of discussions, using open-ended questions to stimulate a wideranging discussion. The interviews were transcribed and coded for detailed analysis, the coding based on themes identified from the discussions, thus reflecting the comments of the interviewees. Analysis was undertaken to examine how the teaching:research nexus was viewed by both managers and academic staff, the factors that influenced these perceptions and how the relationship was managed in practice. In this way, it was hoped to develop working propositions about the factors impacting upon the management of research and how such factors might shape the response of both managers and academic staff. Differences between the four institutions were analysed in order to identify variations arising from the national setting and policy background. In a separate study, the data has also been used to look at the importance of policy matters rather than disciplinary differences in shaping the relationship between teaching and research.

Drivers of the teaching: research nexus

The project revealed a range of key drivers which shaped the commitment to and nature of the teaching:research nexus. These may be characterised as ideological factors, based on the fundamental views and ideas held by individuals or by the universities as corporate bodies, or environmental factors, external forces which impact upon universities and their staff.

Ideological factors

Ideological factors may be defined as those forces that impact upon the relationship between teaching and research drawn from an underpinning body of ideas, beliefs and philosophy. Such factors may or may not be supported by empirical evidence. These beliefs were expressed by both managers and academic staff.

Institutional mission

In some cases, the inter-relationship of teaching and research is expressed explicitly in the University’s mission statement and corporate strategy, and thus formed part of the underlying philosophy of the institution. This was the case with both UK universities. University B asserted that it is ‘‘a research-led institution in which teaching and learning take place in an active research environment’’; University A stated that it would ‘‘.... strive to enhance its position as a leading research and teaching institution cultivating the synergy between teaching and research’’ (my emphasis). Both English universities demonstrated a very strong and public conviction that teaching and research were inter-related. In Sweden, senior officers at both universities argued forcefully that their universities had an equal commitment to teaching and research; both activities were seen as fundamental to the institutional mission. They stressed that Government would not accept any alternative view, and would not countenance an approach which favoured either teaching or research. At the same time, both universities themselves adhered to this view and neither wanted to see any alternative scenario.

Beliefs and values

Senior officers in both Swedish universities had a firm conviction that teaching and research worked together in a strong, mutually beneficial relationship. Beliefs and values were intertwined; not only was there a firm conviction that the relationship existed, but that it was also something of great intrinsic worth. These beliefs and values were deeply entrenched. It was assumed that academic staff would apply their research in their teaching and that teaching would benefit research: ‘‘undergraduates are a greenhouse for new ideas’’ (University C). The relationship was unquestioned and did not need to be amplified or qualified. The precise scope might vary by level of study or by subject, and would be strongest at postgraduate level, but the strength and importance of the relationship was not in doubt. Officers spoke openly of staff needing to be ‘‘up-to-date’’ in order to teach; for some interviewees, ‘‘up-to-date’’ was equated with research activity, but for others a clear distinction was drawn between keeping ‘‘up-to-date’’ and undertaking original research. Officers actively promoted the idea that good research meant good teaching; one senior manager asserted very forcefully that ‘‘the quality of undergraduate programmes can be equated with excellence in research’’ (University C). This view was vehemently held, but when asked to justify this statement with hard evidence, no justification was forthcoming; most significantly, no justification was felt to be necessary. It was not clear that senior management held any firm conceptual or philosophical view of the nature of the teaching: research nexus; the strength and value of the relationship were simply accepted without any clear justification or rationale. In England, interviews with senior officers and managers also showed a complex set of intertwining views and beliefs. There was a clear assumption that the interaction of teaching and research was linked to quality in both teaching and research, and that the active interaction of both activities was a distinguishing feature which marked out their universities from other higher education institutions. Neither university saw teaching and research as an ‘‘either/or’’; both went together. In both English universities, the emphasis was on outstanding researchers who were also outstanding teachers or vice versa; staff who were able to communicate to their students and able to inspire those around them.

Pedagogy

Academic staff in all four universities emphasised the importance of research in ensuring that teaching was up-to-date and relevant to the needs of both students and prospective employers. Many examples were quoted of staff using their personal research in teaching, especially in final year projects and in postgraduate study. Staff stressed the importance of research in ensuring that teaching was strong and informed by latest knowledge. In the four universities studied, there was a strong belief that this could only be achieved by active involvement in research and knowledge creation (as distinct from ‘‘scholarship’’ which was primarily concerned with ‘‘keeping up-to-date’’). Research also encouraged the establishment of high quality facilities and strong groups of postgraduate and postdoctoral staff who enhanced the overall vigour and creative environment of the department. In this way, the teaching: research nexus was both part of the philosophy of teaching and a practical outcome. There was widespread agreement that the nature of the relationship between research and teaching could vary between subject areas and between levels of study. In broad terms, these variations may be summarised as follows:

• Across all subject areas, the pedagogic link was strongest at postgraduate level.

• At undergraduate level, the pedagogic link was relatively weak in the early years for science and engineering, where there was a major body of subject knowledge to be imparted, but the link could be strong by the end of the undergraduate programme when project work was undertaken. An alternative view, but leading to the same conclusion, was that in science and engineering, the pace and complex nature of research did not lend themselves to early years undergraduate teaching.

• At undergraduate level, the pedagogic link could be strong for arts and social sciences in the early as well as latter parts of the degree programme because staff had more freedom to develop course content.

• In many professional subjects (e.g. Law, Medicine and Health professions), the scope for the integration of teaching and research was often limited by the learning requirements of professional bodies. However, in such subjects the use of problem-based learning was also relatively common, an approach seen by some staff as similar to research.

Academic staff recognised that, in practice, the inclination and motivation of individuals were significantly more important than subject area or institutional policy. Opportunities for the integration of research within teaching were identified in all subject areas, at all study levels. Without exception, staff in all four universities were agreed about the benefits arising from the application of research in their teaching: ‘‘more lively teaching’’ (University A), ‘‘more interesting and stimulating’’ (University D), ‘‘staff who teach from their research show more enthusiasm and passion’’ (University C). It is clear that there was a strong ideological commitment to the pedagogic benefits of research-based teaching. There was also a recognition, more weakly expressed, that teaching benefited research. Students could come up with interesting ideas or solutions to problems which would then provide the basis for future research; other staff indicated that they commonly used their teaching, especially at postgraduate level, as a ‘‘testbed’’ for their research ideas. However, in the main, the teaching:research nexus was seen as a one-way process in pedagogic terms, certainly at undergraduate level. A sharp distinction was drawn between the relationship of academic staff with undergraduate and postgraduate taught students, and with postgraduate research students. For postgraduate research students, the supervisory relationship was seen as a genuinely two-way interactive process, linking teaching and research.

Student recruitment

In Sweden, university managers were totally convinced about the importance of research for student recruitment. Recruitment might be seen as a practical concern, more related to the external environment. However, the conviction went deeper and was also part of the beliefs embedded among both managers and academic staff. Prospective students ‘‘know that they are living in a changing world and therefore want to go to a university which is active in research’’ (University D). Recruitment is clearly a very practical concern and heavily influenced by the external environment. Yet the strength of conviction and belief was such that this was also an important ideological foundation for the teaching:research nexus. Senior officers believed that students ‘‘knew’’ that research benefited teaching and that students saw that a strong research profile attracted the best staff. Thus, research was critical in student decision making; weaker universities with little or no research were seen to be struggling to recruit students. Academic staff also stressed the ‘‘professional responsibility’’ of colleagues to share their research with present and prospective students, including the inspiration of future students through contacts with schools.

In England, this ideological commitment was less clear. Some interviewees suggested that prospective students might see staff research as a distraction from teaching. This was felt to be a view promulgated in less research-intensive universities. However, the effective integration of research and teaching was also seen as a positive factor which might assist in student recruitment in an increasingly competitive external environment. In English universities, staff asserted that the interaction of teaching and research was a feature which distinguished their own work or that of their departments from similar departments in other (by implication, ‘‘lesser’’) universities. This was seen as a major factor in the recruitment of high quality students and in producing outstanding graduates on completion of their studies. The view in England, therefore, was that the relationship between teaching and research might be a difficulty in student recruitment or could be shaped and presented to competitive advantage. This was a different view from that prevalent in Sweden where it was assumed that students would expect to see and benefit from the interaction of teaching and research. Recruitment may therefore be seen as both ideological and environmental in its impact. Here, therefore, was a range of ideological beliefs which could influence how universities approached the management of teaching and research. At the same time, it is also possible to identify a variety of environmental factors which underpinned the teaching: research nexus.

Environmental factors

Interviews with staff in the four universities revealed a range of environmental factors which influenced the teaching:research nexus. Environmental factors may be defined as those forces or conditions that impact upon the relationship between teaching and research. In some cases, institutions, or their managers and academic staff, may have no choice as to whether they must respond; in other cases, institutions may have some freedom to accept, modify or resist these external forces. The environmental factors will typically be external to the institution, unrelated to the core beliefs of academic staff or managers.

Assessment and accountability

Academic staff in both English universities all believed that success in research was the key factor in achieving professional progression. This took many different forms: promotion; peer esteem; and financial rewards. From a professional point of view, academic staff maintained a clear distinction between their teaching and research activities. Teaching and research were separately accounted for in terms of time and resource commitments; ‘‘teaching time’’ and ‘‘research time’’ were commonly identified and clearly formed a regular part of the academic dialogue. Examples were also quoted of curricula vitarum which separated teaching and research. Staff believed that they were separately assessed for their teaching and research, in the early years of an appointment, through probation, through subsequent progression and in staff appraisal. In England, there was also a widespread view that teaching was under-valued relative to research. Most staff attributed this to the impact of the Research Assessment Exercise and subsequent ‘‘league tables’’ of universities which had prompted universities to emphasise the importance of research. At the same time, it was also suggested that it was much easier to measure quality and achievement in research then teaching, and that this made research performance easier to assess and compare between staff, departments and universities. The separation of teaching and research in the UK was also emphasised in the external environment. Both English universities stressed the significance for institutional management of teaching quality assessments undertaken by the Quality Assurance Agency and of subsequent institutional review procedures, and of the Research Assessment Exercise; such procedures had a deep and profound effect on the teaching: research nexus, with different arrangements applied to the two elements of the relationship. In England, this separation reflected different requirements for quality assessment and assurance applied by Government, and the perceived absence of any strong Government commitment to the interaction of teaching and research. Both academic and managerial staff referred to the ‘‘assessment culture’’ in England which emphasised the distinction between teaching and research, and which was translated into institutional management practice. In Sweden there was a different fear, namely that the development of more formal assessment procedures might lead to pressures for specialisation or ‘‘division of labour’’ (University C) and thereby threaten the integration of teaching and research.

Market forces

Both academic staff and university officers, especially in England, but also to a lesser degree in Sweden, emphasised the growing impact of market forces and competition on higher education, compelling new forms of management reaction. Fears were expressed that Governments were looking to encourage competition in order to enhance efficiency and student choice; as a result, universities were finding it increasingly difficult to maintain a broad and balanced commitment to both teaching and research. Pressures were growing towards specialisation, in research, in teaching, in student access, in technology transfer and in other priority areas; with greater institutional diversification, it was suggested, went an erosion of the teaching: research nexus. Market forces also required the provision of adequate information for those seeking to ‘‘purchase’’ or utilise teaching and research. This highlighted strengths but also exposed weaknesses, and again tended to emphasise the separation of teaching and research, reducing the scope for a broadbased portfolio of activity.

International and global competition

All the universities studied referred, in particular, to the growth in international competition. ‘‘International status’’ was a goal for each institution. This was normally associated with excellence in research. To this end, the universities concerned were looking to identify key research strengths for particular investment. Both academic staff and university officers recognised that such pressures tended to erode traditional views about the integration of teaching and research and placed a particular responsibility on the institution to manage the teaching: research nexus more effectively. The drive for international standing reflected increasing globa competition, for the best students (especially research students) and for the best staff; senior officers also referred to aspirations of Governments to have universities able to compete on the international stage, especially given the importance of universities in technology development and innovation.

Differential funding arrangements

The research study revealed very different attitudes towards the teaching:research nexus on the part of Government. Both English universities stressed that successive UK governments had questioned the existence of any positive relationship between teaching and research. Most recently, for example, the Government White Paper published in January 2003 commented as follows: ‘‘We believe that the time has come to look carefully at the relationship between research and teaching. In reality, the connection between an institution’s research activity and its teaching is indirect, and there is ample evidence of the highest quality teaching being achieved in circumstances which are not research- intensive. The scale and location of research activity has to be justified and decided on to own merits. We are also determined to promote other sources of recognition, achievement and prestige besides research, both with and between institutions.’’ (White Paper, 2003, para 2.7) Whilst such comments have underpinned moves towards the conferment of the title of ‘university’ on institutions without any research base, it is also true that funding has recently been made available to encourage the development of research-informed teaching, especially in universities that do not receive large amounts of Funding Council income for research. Underlying this initiative, explicit in the documentation, are assumptions about the benefits of research and teaching taking place alongside each other. In the eyes of the two English universities, however, the absence of any Government commitment to the teaching:research nexus has been reflected in differential funding arrangements. Since 1986, English universities have received a block grant from the funding bodies within which resources for teaching and research were separately calculated and separately identified. Both English universities saw these arrangements as a key influence on the relationship between teaching and research. On the one hand, the separation of funding encouraged universities to pursue policies which sought to maximise Government funding from the two distinct streams; many universities also pursued internal resource allocation procedures which were strongly influenced by external income flows. It was also emphasised that there was no financial incentive towards the successful integration of teaching and research. On the other hand, academic staff and university officers stressed that such arrangements underlined the importance of effective internal management actively to recognise, promote and reward the interaction of teaching and research. In Sweden where Government offered more encouragement to the integration of teaching and research, funding of ‘‘education’’ (teaching) and research was also separately identified and universities were required to account separately for funding received. However, staff and officers did not see this separation of funding streams as a major issue or as a key challenge to management. The contrast between the attitudes and response of universities in England and Sweden may be explained by different public and political perceptions. In England, higher education has emerged as a political issue, encouraging Government to intervene more actively and to seek further accountability. By contrast, an officer of the Hogsko¨ leverket summed up the position as follows: ‘‘In Sweden, higher education is not a key political issue; it is not high on the political agenda. Government seems generally content with the quality and relevance of higher education; it is not seen as unduly expensive. People accept the role played by the universities; they do not question it.’’ (Interview) In this atmosphere of compliance, apparent inconsistencies between Government rhetoric and practice relating to the relationship of teaching and research were not seen to be a cause of concern to Government or the reason for a more proactive management response within the institutions.

The response of university management

Having identified a range of ideological and environmental factors influencing the teaching:research nexus at institutional level, it is necessary to consider the response of institutional management. On the basis of the four universities studied, it is possible to distinguish between passive and active management. Passive management may be viewed in essence as a non-interventionist approach, leaving the main responsibility for interpreting and delivering the teaching:research nexus to individual academic staff. Active management involves a more proactive, interventionist approach by institutional management in the development and assessment of the relationship between teaching and research. The contrasting styles of institutional management were themselves the result of differences in the balance between different ideological and environmental factors. The preponderance of ideological factors tended to lead to more passive management of the teaching: research nexus. They provided a set of core beliefs and an essential rationale, but they were non-threatening and non-inquisitorial in nature; they assumed and relied on the independence of academic staff in shaping the nature and interaction of their teaching and research. By contrast, the preponderance of environmental factors tended to lead to more active management. Pressure for increasing assessment, accountability and value for money, and the impact of competition and market forces were driving institutions to specialise in particular areas of activity. Such environmental factors worked in several ways, leading Governments and institutions themselves to separate teaching and research and/or forcing universities to justify, defend and enhance the inter-relationship of teaching and research in a proactive way. Thus, environmental factors tended to contribute to active management within the institutions. In England, in particular, it was clear that university management, as well as individual members of staff, believed that the teaching–research nexus was under threat from a range of environmental factors. Their response was an uneasy compromise between replicating and implementing Government policy at institutional level and a proactive pursuit of the interaction of teaching and research.

Passive management

Passive management of research involved the creation of a supportive but nonintrusive working environment. This should not be equated with a ‘‘do nothing’’ approach; indeed, creating a supportive working environment is a very positive management objective. However, the approach is passive in the sense that direct institutional control over teaching and research is minimal. Characteristics of passive management included the following activities.

Curriculum development

In all four universities studied in the project there was a strong assumption by institution managers, including academic managers such as Deans and Heads of Department, that academic staff would look to apply their research activities in shaping curriculum development, both existing programmes and new courses. For academic staff, this assumption was associated with their assertion of individual freedom to develop their own teaching and research interests; this was viewed as a fundamental aspect of academic life. Universities were keen to support this process, looking to exploit research strengths in the development of teaching programmes. Thus, in one of the Swedish universities studied, new courses had been developed at both undergraduate and postgraduate levels in the fields of visualisation and virtual reality, major research strengths of the university concerned (University C). These developments reflected in part an opportunistic response and a desire to secure a competitive market position, but also, in part, an underlying conviction that research and teaching could not and should not exist in isolation from each other. Both English universities also looked to link teaching and research in undertaking new academic developments, including new lines of research and new degree programmes. Even where particular initiatives were funded by teaching or research resources, both universities looked to develop complementary teaching and research activities. For example, in one of the universities, departments seeking authority for new degree programmes were required to demonstrate the integration of research within teaching and this was specifically monitored in ongoing monitoring and evaluation (University B). The study revealed a wide range of different applications for research in the curriculum, varying from courses based on staff research interests to direct involvement by students within the research process. The universities aimed to create and nurture an atmosphere of enquiry where students would engage with research by asking questions and actively seeking answers. However, the main responsibility for developing the interaction of teaching and research lay with individual members of staff and/or course teams. A characteristic of passive management found in all four universities was the removal of practical obstacles to the teaching:research nexus, leaving the way clear for staff innovation. Examples included the development of organisational structures which encouraged intra- and inter-disciplinary collaborations (including the removal of financial disincentives to the transfer of student load between departments) and the introduction of flexible timetabling with long time periods more suited to in-depth study and research and experimental work.

Stimulation of research All four universities were aware of pressures towards the stimulation of research for reasons of esteem and status relative to peer group competitors. Whilst they were aware of such pressures, both academic staff and university officers emphasised an equal commitment to teaching and research; both were seen as fundamental to the institutional mission. In Sweden, the universities believed that Government shared this view. In England, where the position of Government was different, the two universities were still looking to exploit the link between teaching and research to mutual and overall benefit. In both universities studied, there was strong resistance to the idea of more active intervention and the creation of either ‘‘teaching-only’’ or ‘‘research-only’’ units. In Sweden, in particular, there was a broad acceptance that the stimulation of research activities could benefit teaching. This was not necessarily directed or defined. Rather, there was a general acceptance that resources would be deployed to the overall benefit of all concerned. When staff were able to attract external research funding, it was often possible to redeploy other resources. This could happen at institutional level or within faculties or departments, often to help other strategic priority developments or new subject areas. Some staff expressed concern at such transfers from research to education (teaching), a concern strengthened by a firm conviction that research was under funded; more generally, however, it was acknowledged that facilities developed for research could have a direct benefit on undergraduate students and could broaden their learning experience, especially in science, engineering and medicine. The four universities in the project all had strong research records. However, all four recognised that different members of staff might be at different stages in their research careers and that individual productivity might vary; it was also recognised that different disciplines were at different stages in their research evolution. For example, some health-related studies, such as Physiotherapy and Nursing, were relatively new to higher education and had not yet acquired a strong research base. The universities also recognised that the meaning of research would vary between different staff, disciplines and institutions. Colbeck has demonstrated in her research how one university defined research as the scholarship of inquiry (Boyer, 1990) whereas another viewed research as embracing the scholarships of inquiry, integration, application and teaching. She concluded that ‘‘the broader the university definition of what counts for research, the more faculty are able to integrate research and classroom-oriented teaching’’ (Colbeck, 1998, pp. 660–661). Thus, the definition adopted in each institution is crucial. In the context of the present project, passive management was intimately associated with a broad view of research, a desire to encourage research in whatever form was deemed appropriate and worthwhile by the staff concerned. Active management, by contrast, was more associated with scrutiny and assessment, and tended to encourage a much narrower view of research; the stimulation of research was more often targeted and associated with clear plans and target outcomes.

Quality and relevance

Whilst passive management is, by definition, non-interventionist, all four universities maintained an active oversight of the quality and relevance of both their teaching and research activities. In both Swedish universities, staff were appraised regularly on the basis of both teaching and research. The emphasis was on quality assurance and the enhancement of quality on the basis of formative interaction. Similar procedures existed in the two English universities. However, the emphasis was on assessment against formal targets and performance indicators. In one of the English universities, departmental review procedures specifically addressed the inter-relationship of teaching and research as a holistic process; this was a reflective, formative process, underpinned by a strong belief in the mutual benefits of interaction between teaching and research and by a determination that this be translated into academic and management practice (University D). In both Swedish universities, senior managers emphasised the importance of both teaching and research in securing promotion. It was suggested that, in the past, staff progression had depended primarily on research, but now teaching and research were equally valued; it was not possible to secure promotion on the basis of research alone. The change reflected a growing recognition of the importance of quality and innovation in teaching. Both English universities took account of both teaching and research in considering the initial appointment, probation, progression and promotion of academic staff. Whilst both institutions emphasised the importance of teaching as well as research, many staff took the view that the main factor driving promotion and peer esteem was achievement in research. Significantly, it was not apparent in any of the four universities that the actual integration of teaching and research was a critical factor in the assessment of staff performance; officers in one university were aware of this apparent shortcoming given the institutional commitment to the teaching: research nexus but felt unable to develop effective criteria with which to monitor staff performance (University C).

Active management

Active management meant that the universities to varying degrees sought to intervene in the development of teaching and research in order to ensure compliance with, or fulfilment of, institutional objectives. Those universities which pursued the active management of the teaching:research nexus tended to work in three vital areas of activity (management tools): Strategic and operational planning

The interaction of teaching and research occupied a central, very explicit position within the strategic planning of both English universities. The mission statement of University C refers to ‘‘a research-led institution in which teaching and learning take place in an active research environment’’ and the equivalent document from University D indicated that it would ‘‘strive to enhance its position as a leading research and teaching institution cultivating the synergy between research and teaching.’’ Clearly, both universities felt a need to assert in forceful terms their commitment to the teaching:research nexus as a crucial foundation for their planning activities. Whilst both English universities emphasised the links between teaching and research, in their operational planning they both felt compelled to develop distinct strategies for teaching and research. In part, this reflected external pressures (especially the requirement of the HEFCE for teaching and learning strategies and the importance attached to preparation for the Research Assessment Exercise). In part, it reflected internal management procedures. Both universities had senior officers (Pro or Deputy Vice-Chancellors) designated for teaching and learning or for research, either explicitly titled or holding a recognised ‘‘sphere of influence’’, and had separate administrative departments concerned with, for example, teaching support or teaching quality, or research support in technology transfer and the exploitation of research. It was not clear whether the co-ordination and day-to-day interaction of these operational units reflected the institutional commitment to the integration of teaching and research; in both universities, officers acknowledged that new developments in either teaching or research were often initiated in isolation. Both universities took the view that this did not diminish the overall university commitment, but also conceded that this was an area where there was considerable room for improvement. It was suggested that ‘‘in the end, the relationship between teaching and research is confirmed at departmental level and through the work of individual staff (University C). Thus, active management at university level commonly translated itself into practical separation of teaching and research, and included the positive leadership and direction of both areas of activity; target setting and selectivity in the pursuit of research were both commonplace. The two Swedish universities adopted a very different management approach. One of the Swedish universities (University B) had an overall corporate strategy, which emphasised the importance of both teaching and research. The strategy did not speak in terms of the integration of teaching and research, and did not contain clear targets for either area of activity. Operational planning was left to individual faculties and departments. The University did not possess specific written strategies for research or for teaching and learning at institutional or faculty levels. In the course of interviews with senior managers, it was clear that they felt no need to develop specific institutional plans for teaching and research; instead, their emphasis was on more specific project planning and on the response of individuals to broad guidelines set out at University level. Thus, the University did not advocate the separation of teaching and research in its formal planning procedures. At the same time, however, the University did not actively promote the integration of teaching and research; this was taken for granted and did not require a more interventionist approach. For some senior managers, the explanation, justification, development and planning of the relationship between teaching and research were seen as the business of individual academic staff, an area within which it was inappropriate for the University to intervene; for others, the relationship was simply accepted and did not require further scrutiny. Passive management, therefore, was characterised by a non-interventionist approach on the part of senior managers.

Resource allocation

The two English universities studied had very different resource allocation models, one based on historical expenditure patterns and the other based on income flows. However, in both cases, resource allocation at University level was based on separate funding of teaching and research in academic budgetary groups. Both universities monitored income for teaching and research separately and were acutely aware of changes in income flows and changes between departments, and of comparative information studying the position of departments in other universities. In both universities, resource streams for teaching and research effectively merged again within departments and especially in expenditure on staffing, with considerable final responsibility resting with Heads of Department. Active management of the teaching: research nexus was therefore characterised by separation of income flows and by strenuous efforts to maximise income flows for all areas of activity, to be utilised either separately for teaching and research or jointly within a managed process. Active management was also closely associated with transparency and increasing accountability. In England active steps have been taken through the Transparency Renew and the development of transparent approach to costing (TRAC) methodologies, to unravel the cross-subsidisation between teaching and research funding. Such moves have been stimulated by concerns regarding the under- funding of higher education, but they also compel a proactive approach by university management, especially if the interaction between teaching and research is to be maintained. The universities in Sweden followed a different approach. At university level, there were broad expectations about the commitment of resources for teaching and research which would be varied by income generation, but there was little sense of the competition for funding between institutions and between departments, and between teaching and research which characterised more active management of teaching and research. A key role was exercised by the Head of Department to negotiate with individual staff regarding the balance between teaching and research, with relatively little involvement from senior university managers.

Staff development

The third key activity which characterises active management of teaching and research is a major concern with staff development. Higher education managers emphasised that the delivery of teaching and research relied on the quality of academic staff and therefore needed to be promoted through active staff training and career development programmes. Both English universities had well established policies for the support of newly appointed academic staff, including adjusted workloads and monitoring, and for the continued appraisal of staff; training programmes included many areas of teaching and research. Staff performance in both teaching and research were regularly maintained, although both universities recognised that training and performance tended to emphasise teaching and research as separate rather than integrated activities. Academic staff emphasised the pressures they felt; one academic commented that ‘‘there is no hiding place’’ (University A). Most staff attributed these pressures to the effect of external assessment, especially the Research Assessment Exercise, which prompted universities to manage staff development in a proactive way. Active management placed a strong emphasis on human resource development. The two English universities saw professional training and career rewards as key tools in the stimulation of the teaching: research nexus. It was less clear that current provision reflected these aspirations, but the intention was strongly expressed. There were clear differences in attitudes and practice between the two English universities and the two Swedish universities in the study. In Sweden, staff training programmes were less well developed and staff performance was much less of an issue. In fact, several academic staff expressed concern that the absence of strong management meant that poor performance could be tolerated; the emphasis lay with the Head of Department, commonly a colleague of many years, rather than with University management, leading to acceptance of staff weaknesses.

Conclusion

This paper has attempted to identify the main factors influencing the institutional management of teaching and research. Key concepts have been developed, including the identification of ideological factors and environmental factors and their subsequent translation into passive management and active management, with contrasting approaches to the organisation and character of academic life. These conclusions may be summarised and brought together to form a conceptual model to describe policy and management of the teaching: research nexus (Fig. 1). Teaching and research remain central to higher education. In the present study, among academic staff in all four universities there was a powerful reaffirmation of traditional views of the teaching: research nexus. Staff argued strongly that there were mutual benefits between teaching and research which impacted upon the quality, relevance and delivery of teaching; the positive benefits of teaching on research were less well articulated, but were nonetheless felt to exist. However, the study also revealed very significant differences between institutions in how the relationship between teaching and research was articulated and managed. To summarise, in England, external pressures towards institutional competition, separate assessment and funding of teaching and research, and explicit Government scepticism towards higher education had forced universities to justify and defend their teaching and research activities and the inter-relationship between them, thereby requiring the development of proactive management of teaching and research, and of the teaching: research nexus. By contrast, in Sweden, with a more favourable, sympathetic external environment, the predominant drivers were ideological. As a result, universities felt less requirement to apply the same interventionist management procedures. Active management meant the vigorous involvement at university level of senior officers, including both academic and non-academic managers, in shaping the research activity through planning, funding and staff development. Passive management emphasised the role of individual member of staff in the evolution of their teaching and research, possibly influenced by interaction with colleagues and their Head of Department. In practice, the weighting attached to different factors varied between the four universities studied and will inevitably vary across other institutions. Similarly, the study did not aim to assess the extent to which academic staff sought to link their teaching and research or to establish a grounded theory covering the relationship between institutional management and the actual delivery of teaching and research by academic staff. That would be an important future line of research. However, more modestly, it is suggested that the model proposed in this paper will offer a new conceptual framework within which the management response of individual institutions can be analysed.

The digital transformation is fundamentally changing the way research is done in all fields. On the one hand, researchers are processing and studying more digital data in all research fields, on the other hand, research is increasingly conducted in a networked fashion. The Internet, email, videoconferencing, social media, e-journals, Virtual Research Environments, etc., make it possible to work in large, international, and interdisciplinary teams. The digital transformation of academic research culture poses many challenges concerning the management, protection, preservation and sustainable provision of research data, and the shared use of resources. These challenges can only be solved if all stakeholders, from the individual researchers, to universities, libraries, computing centers, and third-party funders work together. In this sense, we understand the term e-research as the provision and usage of digital tools and data in a collaborative and distributed mode [1]. In Germany educational policy bodies and third-party funders aim to join efforts by giving recommendations [2,3] and guidelines [4] (pp. 21–22), [5,6] on research data management on the national level. The German Wiki Forschungsdaten.org [7] provides up to date information about national and international developments concerning research data management. Currently seven German universities (Bielefeld University, Göttingen University, Heidelberg University, Humboldt-University Berlin, Kiel University, TU Darmstadt, and University ofWuppertal) have an institutional research data management guideline or policy according to a sub site of the above wiki [8]. Given the complexity of achieving sustainable research data management, a task that requires not only special disciplinary and interdisciplinary RDM skills, but also extra human and financial resources, and based on the insight that additional requirements concerning research data management should be accompanied by institutional support [9] (p. 51), [10] (pp. 67–68), it is no surprise that most German universities that have already issued research data guidelines are also setting up an institutional support infrastructure though the timeline, state of development, and approach may vary with individual circumstances. In this case study, we will have a closer look at the situation at Göttingen, where the Georg-August-Universität Göttingen (Göttingen University) has officially set up the Göttingen eResearch Alliance [11] (eRA) as a support infrastructure parallel to the issuing of its research data guidelines [12]. In the following, the Göttingen eResearch Alliance—including its service offerings in addition to those supporting research data management as such—will be introduced in general terms. In the remaining sections, the focus will be mainly on its service elements concerning research data management support to faculty, followed by a discussion and prognosis for the further development of the campus infrastructure.

2. The Göttingen eResearch Alliance

The Göttingen Campus consists of Göttingen University, the University Medical Center, five Max-Planck-Institutes, the Göttingen Academy of Sciences and Humanities, the German Primate Center, and the German Aerospace Center. Together, they form an alliance with a common governance [13]. The members of this alliance and the local infrastructure providers, the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (the computing and IT-competence center of Göttingen University and the Max-Planck-Society, GWDG [14]) and the Göttingen State and University Library (Niedersächsische Staats- und Universitätsbibliothek, SUB [15]), are working in close collaboration to provide solutions and support in the context of e-research activity. Göttingen University, together with the University Medical Center, is one of the first German universities to release a formal research data policy (Research data policy of the Georg-August University Göttingen (incl. UMG), 1 July 2014) [12] to guide its researchers concerning the management of research data. This policy builds on the University’s Principles of Good Scientific Practice [16] and its Open Access Policy [17]. As described by Dierkes and Schmidt [18], this data policy is aligned to the respective national guidelines provided by the German Research Foundation [4–6]. Along with the data policy, the university funds the Göttingen eResearch Alliance with five full-time equivalents of new staff for four years to establish an institutional and centralized infrastructure to support researchers in implementing the policy regarding their regular research practices. Currently, the team of the eRA numbers seven members. Three team members are primarily concerned with data management support, while the others are involved in training, outreach, and service development. This initiative was made possible through the strong support of the university’s Presidential Board, especially the Vice President for Infrastructures. The founding members of the eRA are the SUB and the GWDG and the joint project, which started in October 2014, draws on a long and successful history of cooperation between the SUB and the GWDG. These research-oriented infrastructure providers already have years of experience in e-research related projects and activities and had already successfully cooperated in diverse e-research related projects such as Pericles, GFBio, DARIAH-DE, TextGrid, and the Humanities Data Centre. The upshot of this is that the eRA team can access a resource consisting of some 70 e-research specialists based at the GWDG and SUB. After the initial funding period, the results and achievements of the development project will be evaluated. The Presidential Board, the SUB, and the GWDG will then decide on the details of any long term continuation as an institutional support structure. The project manager and the team are supported by a steering committee, which oversees the overall development (Figure 1). The steering committee consists of the directors and senior staff members of the founding institutions and the project manager (eight members). To promote the embedding of the eRA within the organizational structure of the university and to reflect all major information-infrastructure providers, two of the eight A major milestone in the development of a participatory structure was the constitution of the Göttingen eResearch Council in June 2015. The Göttingen eResearch Council is a Campus network of e-research specialists that provides an interdisciplinary platform for strategic discussions of e-research needs and prospective developments between the Göttingen Campus partners and the eRA. Additional to the Göttingen eResearch Council that will meet twice a year, an international External Advisory Board has been established to evaluate and align the eRA activities on the Göttingen Campus with both other national and international developments. The eRA has made a huge effort during its start-up phase to establish efficient workflows and direct lines of communication, in order to provide individual support to researchers on the Göttingen Campus concerning RDM and other e-research related topics. A crucial element of this consists of strengthening cooperation with the Research Office and other relevant players. We have learned that establishing operational and effective workflows and lines of communication within an extended and complex science hub such as the Göttingen Campus is a complex but critical task. It requires time, patience, diplomatic skills, and subject-specific expertise to bring all stakeholders together. A central component concerning the support of RDM has been the establishment of a workflow to provide subject-specific feedback and assistance on RDM plans for third-party funding proposals, which will be discussed in Section 3. Though the remainder of this article focuses mainly on the RDM activities of the eRA, the full scope of its mission is considerably broader. It was set up primarily to promote e-research on the Göttingen Campus by building a strong e-research network, and to support researchers to fully exploit the potential of the digital medium for their research throughout the whole research life cycle (cf. [19], p. 43). The eRA supports the research life cycle by offering consultation, training, brokering, and project participation (Figure 2). In particular, the term brokering refers to connecting researchers with existing technical resources, tools and methods, as well as with local experts. Project participation refers to collaborative research projects with specific information-infrastructural development needs (cf., Section 3). The website of the eRA [11] plays a central role as an information resource and information hub on e-research for the whole campus. In the following, the focus will be on those measures already introduced concerning RDM support. Based on the experience of the first year of the eRA, insights into how a campus structure for RDM can be further developed and which elements are crucial will be presented.

3. Data Management Support: Development and Implementation

Göttingen University has accepted the challenge of the increasing importance of research data in the scholarly communication process and the implications for universities, and has taken the decision to develop and implement a timely set of guidelines for the handling of research data. The research data policy of Göttingen University and specific funders’ requirements provide the baseline for the support offered. The preparation and evaluation of a DMP is intended to be ameasure of quality assurance. For RDM support to be effective, the central entrance point of contact for RDM consulting is during the grant application phase, during which the eRA can work together with the applicant(s) to develop a data strategy for the bid. In the early stages of the eRA, however, the services were still little known across campus. Larger projects are in any case in contact with the research office, which coordinates the overall application process. The importance of close cooperation with the Research Office is that it helps the eRA-team to become informed about the existence of research projects, as well as ensuring researchers receive information about the offerings of the eRA. This is the principle reason why the university and the eRA have decided to focus on RDM support for larger projects during the start-up phase of the eRA. Gaining intelligence about smaller projects is relatively more challenging. Therefore, the eRA has to actively advertise its services at various levels, such as the faculties and the departments. Another advantage of initially focusing on large projects is that many of the researchers in these consortia are also involved in smaller projects and can act as informal mouthpieces for the eRA’s services. Due to restricted financial and human resources, however, currently the eRA is only able to provide limited support to ongoing or completed research projects in the form of advice, training, and relatively low-level technology developments. The focus on the application process ensures nonetheless that particular informational technological and expert resources for RDM can be taken into account at an early stage. It also has the advantage that an individual data strategy can be developed together with the researchers which allows for the best possible match between scientific and funders’ RDMrequirements, which in turn can improve the overall quality of any proposal. In order to provide tailored RDM support regarding the discipline-specific needs of the individual researcher’s research project, the eRA team is composed of members of different scientific backgrounds.

3.1. Consultation for Third-Party Funding Proposals Especially to the German Science Foundation (DFG)

An important instrument of third-party funding in Germany is that of the large coordinated programs funded by the DFG. These are collaborative research centers (CRCs), research training groups (RTGs) and research units (RUs). Applications for CRCs and RTGs are institutional and the Research Commission of Göttingen University, appointed by the senate of the university, reviews each application before it is submitted to the DFG to ensure that it meets certain quality criteria. Finally, the application is submitted to the DFG on behalf of the university. The overall application process at Göttingen University is coordinated and supported by its Research Office. As discussed above and because such programs are extremely valuable both in terms of kudos and funding, the eRA provides an evaluation of any bid’s data strategy as part of the quality assurance process described above. The evaluation considers the requirements of both the DFG and the Göttingen University data policies and incorporates discipline-specific best practice. As a first step, the eRA had to develop and implement an effective organizational workflow, which involves the researchers, the Research Office, the Research Commission, the eRA, and local IT support. As part of the workflow, the eRA provides an evaluation concerning the data strategy outlined in a proposal to the research commission. To inform the research commission and the researchers about which RDM aspects are taken into account, the eRA has developed a guide in cooperation with the research commission and the research department (Im Qualitätssicherungsprozess betrachtete Aspekte zu Forschungsdatenmanagement und Publikationsstrategie, currently only available in German, [20]). The main topics addressed are the existence of a data management plan, the degree of usage of local IT and information structure, the clarification of roles and responsibilities within the project, the documentation of data and processes, the consideration of human and technical resources specifically for e-research components, the archiving and accessing strategy for the data beyond the project’s life time, and the publication strategy. This workflow is also applied on a smaller scale for DFG proposals for Research Training Groups (RTG) and Research Units (RU). The main focus is on actively and individually supporting and advising researchers on RDM-related issues during the proposal writing process. The quality assurance process is used as leverage for an early consideration of data management issues in an application process. An important aspect of the workflow outlined above is the concept of a consultant who acts as “a single face to the customer”, the researcher(s). It should be noted, however, that in the academic world one should rather think of a partner-ship based cooperation than of a simple provider-client relationship. The consultant is appointed by the project manager of the eRA and selected from among the eRA team and steering committee, depending on the subject area of a proposal as well as its distinctive IT features. She collects the necessary information and requirements related to data management in cooperation with the designated contact person from the research department. In this service-oriented approach, much of the information is gathered ’behind the scenes’ and the number of different contacts each researcher has to handle is reduced as much as possible to optimize the effectiveness of the process. Currently, a set of key questions concerning research data management and publication strategy is being developed as a framework during individual consultations and will be made available on the website of the eRA. These questions are based on authoritative sources, such as checklists provided by the DCC [21] or theWissGrid-Projekt [22], and refer to the guide for statements provided by the eRA (see above). In certain cases, the DFG provides support for data management and other e-research related developments within large coordinated research programs. This opens up the opportunity to fund researchers, data managers, and IT-support technical staff. Regarding CRCs, these developments are possible as INF (Information Infrastructure) sub-projects (cf., [23], esp. p. 114). Through such cooperation it is possible to implement sustainable research data management and e-research methods at the center of such large programs. This ensures that relevant issues are addressed during all stages of the research data life cycle and that adequate resources that meet infrastructural needs are planned from the beginning. The concept of embedded data management is relatively new and still evolving. The skills of a data manager appear to be a mixture of those of an IT and research-domain specialist (cf., [24]). Embedded data management seems to offer an appropriate solution for the requirements of large projects with diverse and complex data. It ensures support for the handling of research data along all project phases. The acceptance of the concept of the INF-projects and embedded data managers or data librarians requires a change in research culture, as ideally the data managers are cooperatively embedded into the research projects. Such a departure will require time to become standard research practice (cf. [23,25]). It is part of the work of the eRA to identify at an early stage needs for project-related infrastructure developments and to support the process for applying for funding. The eRA is also associated to some infrastructure projects. In those projects without a full-scale INF sub-project, or without permanent support positions, the eRA is exploring alternative ways (backed up by the human resources of the two founding institutions, SUB and GWDG) of supporting such researchers, for example by providing a combination of eRA-resource pooling and intense consultation and training.

3.2. Further Elements of an e-Research Support Infrastructure RDM support is not intended to be limited to the application process. Other elements are providing training on data management practices, developing specific technical services for researchers (e.g., for the publication of data), and participating in research projects.

3.2.1. Face to Face Training’s, Online Training, and Information Materials The area of RDM training is one of the eRA’s fastest growing areas of activity. The eRA provides on-demand training for early career researchers in large coordinated research programs and other institutions on the Göttingen Campus. There are also open workshops for researchers and library staff. RDM training activities vary in length (from 15 min up to two days) and topics may include RDM basics, Open Data and Open Access basics (in cooperation with the department Electronic Publishing of SUB), and can be in the form of hands-on sessions on the use of specific IT-services available on the Göttingen Campus (e.g., cloud storage, research databases). For the training of early career researchers to be tailored to the specific needs of an individual group, the eRA and the coordinator of the group agree upon the content and format of the training program in advance. If necessary and appropriate, the eRA will call in an expert for a specific topic. It has proven especially fruitful to directly involve the participants into the preparation phase. This can be done, for example, by interviewing volunteers about their current research data practices and their specific needs and expectations, and by actively involving them in the training as presenters of specific use cases. Although this approach seems, as far as our experience goes, to increase the effective reception of training on the part of those attending, their direct involvement is neither always possible nor desired, since it places further demands on their (little or nonexistent) time. Especially PhD-students connected to large collaborative projects are mostly from very different disciplines. Therefore, it is rather difficult to give specific recommendations concerning tools, formats, and retention periods for data to such multi-disciplinary groups. For these groups, the training focuses on RDM basics in order to raise general awareness of the value of data in the research process, the Research Data Policy of Göttingen University and its implications, the handling of data in their research group (if applicable) and the requirements of other funders. This includes the definition of a research data strategy, an overview of the available infrastructure for storing data, documentation, and sharing of data, as well as several related aspects such as backup solutions that are available on the Göttingen Campus, metadata, and licenses. Further possibilities are collaborations with other initiatives and institutions on the campus to tailor discipline specific training, such as for the Humanities (together with DARIAH-DE and TextGrid) or the University Medical Department (together with the Institute for Medical Computer Science and the Institute for Medical Statistics). Hitherto, groups of young researchers who participated in eRA-trainings came from quite diverse areas such as (agricultural) economics, interdisciplinary statistics, cardiovascular science, and chemistry. The training for chemists was part of a new Information literacy menu for research groups of the Faculties of Chemistry and of Physics [26]. In this context, the eRA will provide on request pre-defined training blocks ranging from 15 up to 90 min on topics such as RDM or IT-services on the Göttingen Campus. One of the emerging trends is the growing involvement of research libraries in the support and consultation of researchers during the digital research process, i.e., as infrastructure partners or by bringing library employees on a temporary basis as embedded data managers or data librarians into collaborative projects. As a result, there is a need for librarians and library staff to stay up to date with recent developments in this field and to acquire complementary skills as described for example in Engelhardt [23] (esp. pp. 107–108) or Akers et al. [27]. The eRA has started to address this demand by providing two advanced training courses on RDM that are open to all librarians and library staff in cooperation the Service Zentrum Hochschulbibliotheken of the SUB. These courses were well attended by participants from all over Germany, which emphasizes the interest in this topic. These courses also provided the eRA with new insights into the potential for developing user-oriented e-research services and the possible roles for research libraries and computing centers based on their specific competences. The content of these courses was based on the general RDM basics for PhD students mentioned above, but with more focus on institutional approaches and the potential roles for libraries. Given the very mixed backgrounds of the participants and the diverse development of RDM at German universities and other institutions, it proved very difficult to serve all needs of the participants, especially those already very advanced in the topic. From this experience we learned that there is definitely an audience for a RDM basic course for librarians, but also a need for a specialist discussion forum as well as room for dedicated sessions on institutional RDM support alongside library conferences or other related conferences or workshops. Besides face-to-face training and slots, the eRA also strives to provide comprehensive training and information materials on e-research through its website [11]. Therefore, over the past year, the content of the eRA-website has been significantly extended. The focus is on providing general information on e-research combined with subject specific examples, local IT and online services and contact with experts. Currently, a database of local research facilities is also under development. Another essential feature of the new website is a news section in blog format. Though this is not a classi training format, an up-to-date blog post on e-research relevant events, services, methods, projects, or researchers can offer the opportunity to provide poignant in-depth information and is a powerful communication channel to reach more researchers and other relevant players on the Göttingen Campus and beyond.

3.2.2. Participation in Projects

There is a considerable demand on the Göttingen Campus for small, limited services like adaptations of software systems, the maintenance of websites or wiki platforms, or support for data curation. However, these services are not free of charge, as they require expert human resources. The eRA is currently investigating the possibilities of using resources already available within the SUB and GWDG to meet this demand. Another path currently being tried out is to employ student assistants for tasks which have a potential for wider application on the campus. Additionally, the eRA helps projects in need of more extensive support to draft a proposal, e.g., to the DFG, to apply for funding for designing and implementing appropriate solutions. In most cases, such personnel will comprise domain experts. Through such initiatives it might be possible to establish domain-specific “satellites” in the individual faculties. One such example is the embedded data manager as explained above. The eRA is currently associated with two such projects in the environmental and in the physical sciences. In both projects, the data manager is funded by the CRC on a 50% basis in a position lasting four years. In one case, another 50% is related to training and consultation services. The two projects differ, in that for the project on environmental research, the data manager is truly embedded into the research workflow. The data manager of the other project is much more concerned with development of software and technical infrastructure, i.e., is more on the traditional support side. Data management in large coordinated programs requires a considerable institutional commitment. The infrastructure, which is developed within such projects, as well as the experience gained, are of great value for the further development of the campus infrastructure in general.

3.2.3. Publication of Data

Part of the scientific process is the publication of results. This often happens in textual form, e.g., as a journal article or a monograph. According to the DFG’s memorandum on Safeguarding Good Scientific Practice [4] (pp. 21–22) and the DFG’s Research Data Guideline [6], the data underlying the publication should be stored or archived for at least ten years at the institution of origin or a trusted repository. It is also desirable to publish the data, thereby simplifying reproducibility and enabling reuse. Data can be archived at the GWDG or at a subject-specific or general purpose data repository (which can be identified for example via the Registry of Research Data Repositories (re3data, [28]). These solutions involve costs and trust in the quality and longevity of the repository service. There are indications that the recommendation to store data for 10 years is an ideal not yet practiced fully by researchers. This is significant in the context of further work indicating that many researchers require even greater longevity of support for their data. For instance, the RMD survey of the Humboldt-University revealed that only 56% of all respondents keep their data for 10 years or longer ([9], p. 21), leaving a lot of room for improvement on the part of institutional research data service providers to make storing data more straightforward for researchers. Although no local survey concerning RDM has been conducted to estimate the concrete practices at the GC, it can be supposed that the situation is comparable. One of the most promising measures to improve the storage and publication of research data might be the provision of a central local institutional data repository [9,29,30]. In this context, a repository which is maintained by a robust institution with a natural long-term view of information management offers advantages. Universities as enduring institutions can maintain trustworthy archives and repositories and facilitate access to storage solutions with local support. Göttingen University already offers central institutional services for the electronic publications produced by members of the university (GoeScholar, GoeDoc, eDiss, Göttingen University Press). The department of Electronic Publishing of the SUB coordinates these services. A central institutional solution for the associated research data underpinning these publications is still however missing. Therefore, the eRA is currently coordinating the development of a research data repository for the Göttingen Campus. Based on LibreCat, this repository will soon provide a home especially for long-tail research data, for which no subject-specific repository exists. The repository will be connected to the current research information and publication management system of the university, which is currently being upgraded. Additionally, the SUB is developing a service to register Digital Object Identifiers (DOIs) in the humanities subject area in its role, since 2015, as the DataCite member in Germany with a particular focus on the digital humanities, as indicated on the Data Cite Website [31]. This component will be integrated in the local repository to allow the discovery and citation of data, as well as cross references for all disciplines. The eRA will act as a broker for the DOI service on the Göttingen Campus.

4. Conclusions and Outlook

Although most German funders and research policy makers recognize the importance of RDM, there are currently no compulsory RDM policies in place. This might in part be due to the constitutional protection for freedom of scientific research (article 5 of the German Constitution) which makes regulations of the research process a politically sensitive topic. Based on the insight that sustainable RDM will only become (voluntarily) an integral part of the research process if its support is considered as an institutional objective, Göttingen University has instituted a local research data guideline and established the eRA as an appropriate institutional support structure for its implementation. The activities of the eRA are aimed at providing an optimal information infrastructure for research by bundling knowledge about existing IT and information scientific resources, solutions, and experts, and acting as a central point of contact concerning e-research. Joining up the diverse fields of activities, as well as the professional expertise and excellent technical resources of a computing center and a research library is an important step to build comprehensive, coherent, and sustainable institutional support for RDM and fostering data-driven research on the Campus. As part of this effort, the seven-person, multidisciplinary team of the eRA provides general support and training for RDM. An alliance of strong infrastructure partners, such as those represented in the eRA, providing institutional support for RDM and e-research offers great opportunities for promoting such research. When allocating special funds to the development of an e-research infrastructure for the Göttingen Campus, both the library and the computing center have committed themselves to fund permanent positions for e-research tasks after the initial phase (2014–2018). The details of the duties and the job specifications will be subject to an evaluation within the University towards the end of the project phase, taking the recommendations of the External Advisory Board into consideration. Implementing new communication structures and workflows requires time, since it involves the adaptation of established workflows throughout the Göttingen Campus and the establishment of new lines of contact between the eRA and the individual players on the Campus. The process involves top-down and bottom-up components, such as quality assurance and the articulation of specific requirements by researchers. After having established the RDM quality assurance workflow as part of the internal review process of DFG grant applications, the initial reluctance (due to the top-down approach) is gradually giving way to a more open and cooperative atmosphere during the support process. Among the lessons learned is making young researchers aware of the importance of RDM. They are the vanguards of the digital transformation and of the transformation of individual research cultures and can serve as multipliers in communicating the importance of RDM into the faculties and individual working groups. Since the end of 2014, the eRA has initiated support for about 25 projects of various types (CRC, RTG, and RU) across all disciplines. Currently, considerable efforts are being invested in a consultation process that is based on the evaluation of individual project bids (on the average about one week per project, depending on the complexity of the project and the familiarity of the researchers with the e-research related aspects of their project). However, it is envisaged that in the course of time the expertise that the eRA is gaining in this way will be translated into more generic case studies. These can be used to efficiently scale up this kind of research support to the Campus. Consequently, more attention can then be given to the remaining advanced cases. The eRA will additionally make more generic training courses available via its website, which will provide extensive information on e-research related topics for individual consultation by researchers. Success of an initiative such as the eRA strongly depends on its acceptance throughout the Campus. In particular, tangible services have to be offered to researchers. There are two levels of engagement involved: (i) the gathering of support from the departments of the central information infrastructure providers, as well as by on-site IT and data-management units; and (ii) the design of products and services that neatly fit into the researchers’ workflows. Personal contacts, especially during the build-up phase of the initiative, are extremely important for the dissemination of information. The communication operates in several directions, for example between the eRA and the researchers. While promoting service offerings, the eRA receives feedback and requirements from the researchers regarding service development. Communication between the eRA and other players (e.g., central administration) will support developing policies and establishing effective administrative workflows. A further consideration in this context is that, since the eRA currently only has the status of a project, related staff turnover precisely carries the risk of losing contacts and knowledge. During the last months the eRA has developed a core service portfolio and focuses now on concrete online services for researchers on the Campus (e.g., data repository, research facilities portal, DOI allocation). Still, it is essential to continuously question the demands of the researchers to improve the eRA-services and to identify new needs. This is partly done by analyzing the consultations and training conducted by the eRA and will be elaborated by carrying out expert interviews. Acknowledgments: The authors kindly thank the anonymous referees for fruitful comments and suggestions. Special thanks go to Mike Mertens (Göttingen University) for additional comments and editing suggestions. The Göttingen eResearch Alliance is a project of the Georg-August University, Göttingen and is run mutually by Niedersächsische Staats- und Universitätsbibliothek Göttingen and Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen. Author Contributions: Both authors, Jens Dierkes and UlrikeWuttke, are members of the eRA project, respectively have been members at the time of the writing of this article (UW). As such, they are involved in all aspects of the program of the eRA and are both contributing equally to this article. Conflicts of Interest: The authors declare no conflict of interest.

In this article, we argue that an engaged pedagogy of research writing forms a critical resource for human geographers navigating the fraught institutional politics1 of research writing in our everyday lives. Motivating our intervention is a concern that the sources, implications and impacts of mounting pressures on our research writing are far from benign. While the lived experiences of human geographers are inherently diverse,2 common (albeit always unstable) ground emerges around shared experiences of escalating demands and ‘metric dogma’ pursued by university managers (Larner and Le Heron, 2005; Rogers et al., 2014; SIGJ2 Writing Collective, 2012). Enmeshed in speeded-up systems, many academics dutifully press ahead: publishing articles, meeting dead- lines, teaching students, writing grant applications and working longer hours. Because of the above pressures, we often default to past approaches of ‘learning-by-doing’ when super- vising PhD students and supporting colleagues (Bathelt and Gibson, 2015; Dufty-Jones, 2018). The hope is that, as in the previous eras, immersion will be enough to launch and sustain scholarly careers. Most worryingly, pressures brought to bear on the production and value of research writing are jeopardizing personal and community relationships and enflaming a crisis of mental and physical health in the academy (Berg et al., 2016: 172; Mountz et al., 2015; Mullings et al., 2016; SIGJ2 Writing Collective, 2012). Such matters have culminated amidst the present COVID-19 coronavirus pandemic, as we grapple with research ambitions while shifting to online teaching delivery and juggling work-from-home alongside diverse caring responsibilities (circumstances we will reflect on at the end of our article). With the wellbeing of ourselves, colleagues, students and the discipline at stake, we have reached a critical juncture. Responding to these challenges, we advance an approach that emphasizes the social, temporal and spatial dimensions of learning the craft of scholarly writing: an engaged pedagogy of research writing. Our approach challenges the often taken-for-granted institutional practices that position individualized, competitive and metricized technologies as the best means of achieving excellence in research writing. An engaged pedagogy of research writing aims to replace these practices with collectivized understandings and strategies that are designed to support the continual honing of our research writing ‘craft’. At its core, our approach prioritizes an ethic-of-care and argues that scholarly excellence in our research writing is more productively achieved through the care-full cultivation of a range of interdependent relationships, times and spaces. 1 Engaged pedagogies of research writing In advocating for an engaged pedagogy of research writing, we understand that everincreasing expectations of academics are refracted by geographical differences and biases, marginalizations and inequalities, and the corporatization and neoliberalization of universities (Berg and Kearns, 1998; Larner and Le Heron, 2005; Mott and Cockayne, 2017; Rogers et al., 2014).3 We also acknowledge that there has always been an institutional politics to academic writing (Bourdieu, 1988) and recognize that it would be misguided to think there ever was a ‘golden era’ when research writing was produced in ideal circumstances (Mendick, 2016). Yet, we also believe that there is something qualitatively different in how the politics of research writing manifests in contemporary universities, especially the Anglophone academy with which we are familiar. For instance, paradoxically, we are offered expanding opportunities for our texts in terms of outlets and styles,4 while the same possibilities are increasingly constrained by what ‘counts’ as research writing. Academic work is both diversifying and extensively codified and surveyed. Our research writing has become structured as a value-generating labor task undertaken as workers within a (proto-capitalist) labor process (Berg et al., 2016). Furthermore, long-held practices and values of scholarly writing – such as rigour, excellence and peer review – have become colonized and are reflexively used as a means of generating a ‘maximum return’ to universities via a system of elevated expectations, growing surveillance and eroding rights (Davies et al., 2021). Our relationship to research writing has become both more calculative and confounding. We regularly find ourselves critiquing the systems and practices that seek to categorize, rank and relationally establish hierarchy in our writing, while also continuing to use and reproduce those systemic structures, values and performances (including, we admit, ourselves writing for this august journal). These shifting, often neoliberalized,5 governmentalities are what we mean by the institutional politics of research writing pervading our daily working lives. Nevertheless, we are also inspired by various interventions developed over a number of decades by feminist and anti-racist scholars. Such contributions have provided a suite of radical interpretations and approaches to draw upon when confronted by changing and mounting pressures (Hawkins, 2019; Mountz et al., 2015; Puawai Collective, 2019). Common across these is an ‘ethic of care’ (Lawson, 2007; 2009). That is, a shared aim to move beyond critique, forging new ways of writing and foregrounding relationships and practices that prioritize and enhance care for self and others. Furthermore, ‘starting with a lens on care brings to light not only the resilience of care but also the transformative potential of care ethics in contexts undergoing reform’ (Power and Bergan, 2019: 433). Higher education is one such context from which care, as an ethic, can form the grounds for resistance. The task is to incorporate care into professional practices of scholarly writing to transform the institutional values, policies and practices that structure our writing. However, given the intensification of metricdriven university management, we argue that the current suite of ‘care-full’ strategies developed thus far is insufficient. Specifically, we seek to address a problematic disconnection between, on the one hand, contemporary ‘care-full’ interventions around the institutional politics of research writing and, on the other, pedagogy. This gap is already well known. As Castree (2003: 283) observed, discussions in our discipline are marked by ‘a conspicuous non-debate over pedagogy’. Indeed, pedagogy is peripheralized in the academy more generally – something conscious to us when submitting an article with pedagogy at its heart to a research journal with the disciplinary standing of Progress. Notwithstanding these misgivings, our aim is to open discipline-wide dialogue on how existing ‘care’ strategies respond to the institutional politics of research writing, how they can be augmented, and whether new approaches can be developed to enact change that is both generative and enduring. We contend that a pedagogical focus, namely one that combines the insights of both an ethic of care and critical pedagogies of research writing, offers one means to this. In approaching the politics of research writing from a critical pedagogical perspective, we recognize that pedagogy and the politics of research writing are intimately and iteratively linked. While all aspects of this process compel critical reflection, our article consciously turns the lens of critical pedagogies to focus firmly on the politics of research writing within our own spaces of institutional and professional practice. By focusing on the institutional politics of research writing, we draw attention to how ‘[w]hen things become institutional they recede’ from our consciousness (Ahmed, 2012: 21). Nowhere is this kind of institutional ‘taken-for-grantedness’ more evident that in the problematic disconnection that exists between the routine activities of teaching and learning and the practices of research writing. Yet, by attending to ordinary features of institutional life and making explicit connections between pedagogy, research writing and the politics therein, we believe that opportunities are generated for human geographers to not only work at our institutions but also to work on them (Ahmed, 2012). This article proceeds by reviewing how human geographers have responded critically to the institutional politics of research writing through an ethics of care, noting the absence of pedagogy in such responses. In the third section, these geographical insights are brought into dialogue with the ‘pedagogies of research writing’ and ‘engaged pedagogies’ literature. Combining the insights of these three literature – and drawing to the surface the ethics of care implicit in all – we propose an engaged pedagogy of research writing as a concept for the critical and productive engagement with the institutional politics of research writing. Elaborating on this, we propose three practical features of future engaged pedagogies of research writing: (1) understanding writing as social practice; (2) reimaging the temporalities of writing and (3) reimagining the spatialities of writing.

II Responding to the institutional politics of research writing via an ethic of care Our article begins with and is founded on the extensive critiques advanced over a number of decades by feminist and antiracist geographers regarding the dynamics of knowledge production (Bondi, 1997; Monk et al., 2003; Mott and Roberts, 2014; Noxolo, 2009; Peake and Kobayashi, 2002; Rose, 1993). Interventions developed and employed to challenge the politics of research writing include writerly collaborations, non-traditional ways of ‘writing-and- citing’ research, and the different way of constructing research writing ‘markets’. Human geographers’ writerly collaborations illuminate the multitude of relationships that contribute to and sustain our scholarly writing (Mountz et al., 2015; Pratt, 2010; SIGJ2 Writing Collective, 2012; University of Kentucky Critical Pedagogy Working Group (UKCPWG) et al., 2015). Such collaborations challenge the (neoliberalized) notion that success in research writing is the product of autonomous individual efforts. Gibson-Graham (1996: xi) argued that the creation of a singular writing persona allowed them/her to ‘subvert in a practical fashion the myriad hierarchies of value and power that (in shifting and complex ways) structured our relationship’. Wright et al. (2007: 155) and Bawaka Country et al. (2015) rejected the approach to research (and writing) that constructed both the research and the resulting texts as an ‘individual, isolated [and human] task’ pursuing instead a series of writing collaborations that allowed them to ‘commit to long-term research relationships’ with each other and their research collaborators. This includes the Aboriginal land (Bawaka Country) that actively provided nourishment to themselves and their research. Collaborative approaches focus attention onto how the relationship between writing and research practices perform an ‘ontological politics’ (Fisher et al., 2015: 19; see also Barnes and Duncan, 1992; Dufty-Jones, 2020; Eshun and Madge, 2012). Our academic writing not only represents; it also constitutes and brings ‘worlds into being’. In a similar vein, creative approaches to how geographic knowledge is expressed and referenced – how we ‘write-and-cite’ – have also been used to challenge what constitutes an ‘academic’ text (e.g. poetry, song and art), publication spaces, peer review processes and grant application assessments (Hawkins, 2021). Such interventions question what is ‘success’ and how it should be represented and evaluated in relation to our research writing (DeLyser and Hawkins, 2014; Madge, 2014; Mott and Cockayne, 2017; Thomson, 2018). Such approaches creatively reassemble and reflexively employ traditional research writing practices to resist various politics of scholarly writing, making room for authors and expressions that are frequently marginalized and ignored. Meanwhile, our research writing has the potential to be used strategically to resist and reimagine how this work is performed, produced and valued. While our writing outputs are increasingly metricized, ‘market’ characteristics also open up possibilities for making ‘a bid for what [research writing commodities and their markets] might become’ (Smith, 2005: 1). The constructed nature of academic publishing markets generates opportunities for questioning their ‘essence’ and reimagining and remaking them through an ‘ethic of care’ (Smith, 2005). An example arises from digitally disruptive academic writing and publishing efforts that circumvent institutional gatekeepers to gain access to both traditional and new audiences. Digital technologies have facilitated the development of e-journals (e.g. ACME: An International e-Journal for Critical Geographers) with open subscription and multiple languages to make content more widely accessible. Blogs and social media have been similarly adopted to innovate collaborative writing practices and dialogues.6 Ligamented by an ‘ethic of care’, the above approaches challenge the institutional politics of research writing by shining a critical light on the ‘win-at-all-costs’ academic cultures that valorize and reward academic practices of individualization and competition (Davies et al., 2021; McDowell, 2004). As Lawson (2007: 1) notes, care ethics also allow such interventions to move ‘beyond critique and towards the construction of new forms of relationships, institutions and action that enhance mutuality and well-being’. An ethic of care therefore provides a foundation built upon ‘social relationships of mutuality and trust (rather than dependence)’ in our research writing (Lawson, 2007: 3). Such interventions become an end, rather than the means (Atkinson et al., 2011), from which to resist and rework the institutional politics of research writing. Nevertheless, we contend that the aforementioned strategies, as they stand, remain insufficient to achieving the kind of paradigmatic shift urgently needed. Human geographers are yet to advance the possibilities of pedagogy in identifying problematic disciplinary practices around scholarly writing while strengthening ‘carefull’ interventions. Our interest is to establish such a connection. As we next outline, the possibilities of pedagogy can be pursued through the intersection of two distinct literature: (1) pedagogies of research writing and (2) engaged pedagogies. III Bringing a pedagogical perspective to the institutional politics of research writing 1 Pedagogies of research writing As doctoral student numbers increased globally during the 1990s and 2000s, a series of critiques also emerged regarding how academic writing was taught (Aitchison and Lee, 2006; Kamler and Thomson, 2006). The consensus was of a general neglect. Furthermore, when research writing was taught, approaches were ad hoc, or were framed in ‘positivist’ terms with scholarly writing understood as linear and singular in meaning. Likewise, research writing problems were regularly individualized and framed as a personal failing. Solutions were therefore approached as a simple matter of addressing a deficit in skills. In response to such critiques, pedagogies of research writing were advocated as a means of radically changing how to approach the teaching and learning of research writing. There have been two main thrusts: first, writing should be seen as a social practice and as an integral part of the identity work involved in the ongoing (re)production of the academic researcher (Kamler and Thomson, 2004; Murray, 2014). Understood as inherently relational, our writing not only represents but is also conditioned by and productive of social relations and practices both within and outside of the academy. The second thrust was to question the construction of academic writing as static and linear – as something that occurred substantially in the final stage of the research process. Instead, it was pointed out that meaning was often elusive and that writing helped to bring ideas into being (Clughen and Hardy, 2012; Lowenthal and Wason, 1981; Richardson, 2003). Scholarly writing was therefore necessarily messy, iterative and recursive (Cameron et al., 2009). In order for the productive and open-ended capacities of writing to be included in every stage of a research project, the temporalities of research writing needed to be reimagined. We will revisit these strategies in the next section. Beforehand, we wish to point to three aspects of the pedagogies of research writing literature that warrant critique. First is the unidirectional approach to learning often implicit in this literature. That is, pedagogies of research writing approaches tend to focus on what the teacher can impart to the student, neglecting to account for how those teaching research writing are also always learning through their interactions with students, colleagues and communities (Murray, 2014). Second, while advocates for the pedagogies of research writing acknowledge the social and temporal messiness of learning and doing research writing, there is a notable lack of critical engagement with the role of space – understood as the constitutive material and relational spaces within which the writing labor process occurs (Kamler and Thomson, 2004). As geographers, we find this especially problematic. Last, while research writing pedagogies recognize factors brought to bear on the experiences of teaching and learning research writing in the modern university (Aitchison and Lee, 2006), they often fail to acknowledge and confront the institutional politics that increasingly engulf how we and others teach, learn and practice our scholarly writing. 2 Engaged pedagogies In responding to the above problems, we see considerable value in introducing critical pedagogies to the conceptual mix.7 Described as ‘what happens when critical theory meets education’ (Morgan, 2000: 274), critical pedagogies seek to account for the role of human agency in the production of knowledge, recognizing the classroom as a site from which relationships of power can be produced and reproduced both within and beyond the sphere of the school or university (Freire, 1970; Giroux, 1993; Hooks, 1994; 2003). Just like scholarly writing, the pedagogical is inescapably political (Giroux, 1993; McLaren, 1998). Of particular relevance is the way critical pedagogies have been reflexively employed by human geographers to examine the politics of academic working lives and spaces (e.g. Laliberte et al., 2017; Mott et al., 2015; Newstead, 2009; SIGJ2 Writing Collective, 2012; UKCPWG et al., 2015). As Laliberte et al. (2017: 36) argue, critical pedagogies can be used as tools to ‘question and redefine long-standing beliefs and practices in education’. This is achieved both in terms of how we approach our own classrooms but also in terms of how these perspectives can result in transformations of the academic self. This transformative potential is at the centre of a specific strand of the critical pedagogies literature: engaged pedagogies. Coined by bell hooks (1994), engaged pedagogies build on critical approaches by placing an emphasis on care.8 Specifically, ‘an engaged pedagogy contests the center from which care and responsibility can be talked about and experienced’ (Madge et al., 2009: 43). Seeking to disrupt the traditional, binarized and unidirectional teacher–student relationship, engaged pedagogies require academics to be ‘actively committed to a process of self-actualization that promotes their own well-being’ (hooks, 1994: 15; see also Stratford, 2007). For hooks (1994), academics need to recognize and be open to how pedagogy has a transformative impact on the academic self. In particular, we need to bring ‘a willingness [and expectation] to engage, challenge and be challenged’ to our pedagogical relationships (Newstead, 2009: 81). In this way, we remake ourselves through interactions with others, including our students (Madge et al., 2009; Newstead, 2009). Viewing research writing pedagogies via the lens of ‘an active politics of becoming’ (Connelly, 2002; Gibson, 2001) identifies opportunities to break with seemingly hegemonic power relations by creating new care-full possibilities. Thus, while pedagogy traditionally compels a focus on students and their learning, an engaged pedagogy insists that we must also take care of the academic self. This is not self-indulgence. Rather, as Stratford (2007: 175) observed, ‘self-care is crucial to survive in academic life’. Moreover, ‘[t]here is a politics to exhaustion. Feeling depleted can be a measure of just what we are up against’ (Ahmed, 2013). Caring for the scholarly self could therefore even constitute an ‘act of political warfare’ (Lorde, 1988: 131): protecting, preserving and focusing our energies so that we can continue to teach, learn and write. Therefore, if we are to effectively cultivate current and future generations of writerly subjects, we also need to be committed to practices that promote our own well-being. This is an important difference from other pedagogical approaches to writing that tend to narrow the focus onto students and their learning. The implication is that academics have nothing more to learn and/or that we are also not in need of care when it comes to our scholarly writing. Nevertheless, those writing from critical and engaged pedagogical perspectives have, for the most part, neglected to consider how this work could inform how we understand, teach and learn about research writing (exceptions include Giroux, 1993 and Heyman, 2004). As shown in Figure 1, each of the three discrete literature engage with different aspects of the politics and pedagogies of research writing but, to date, they remain problematically disconnected. We propose the development of engaged pedagogies of research writing as a strategic intervention to address these limitations. By establishing hitherto neglected connections between the above literatures, engaged pedagogies of research writing make available new ways in which human geographers’ can both augment and sustain our contemporary efforts to challenge and transform the politics of research writing.

IV Engaged pedagogies of research writing: from ‘homo academicus’ to ‘discipuli curans’ The engaged pedagogies of research writing for which we advocate combine the different insights and strengths of all three approaches outlined earlier in this article (see Table 1). In forging this tripartite connection, engaged pedagogical approaches to research writing aim to build on and extend this work. Engaged pedagogies of research writing prioritize the critical curation (i.e. sharing, teaching and learning) of scholarly writing practices. As different approaches to doing research writing are shared, learned and taught, they are also actively reworked and reassembled in response to specific social, temporal and material circumstances. To that end, in the remainder of this article, we identify three practical strategies for how future engaged pedagogies of research writing could be practiced: (1) understanding research writing as social practice; (2) reimagining the temporalities of research writing; and, in concert with these, (3) reimagining the spatialities of research writing.9 The following approaches are not ‘blueprints’ or ‘best practices’ (hooks, 1994: 10). Rather, they are tools to reflect on how engaged pedagogical perspectives to research writing generate new critical perspectives and practices (that are always contextual and unfinished). To engage with and resist the various institutional politics of research writing in our everyday academic lives will require diverse engaged pedagogies; ours are a starting point. 1 Understanding research writing as social practice A variety of relationships and practices become visible when we understand research writing as a social process. Cameron et al. (2009) describe this as the ‘betweenness’ of writing; it can include social relationships between student and supervisor, collaborators, or individuals in a writing group. Social aspects of research writing can be built more firmly into scholarly writing practices. One example is ‘curating’ who is brought in to review draft texts (e.g. self and peer assessment, writing groups) and how to provide feedback on writing intended to constructively build confidence and capacity. Similarly, writing can be progressed through socially oriented practices such as weekly ‘shut-up-and- write’ sessions (Dowling et al., 2012; Ferguson, 2009; Pain and Mowl, 1996). Chris conducts regular ‘Writers Anonymous’ sessions with graduate students (loosely based on the format of Alcoholics Anonymous) to share ongoing frustrations and challenges with writing in a supportive environment. The format has evolved in iteration with students’ participation and remains very open, with no agenda, seeking to break down barriers and demystify the writing process rather than confer skills. All social interactions around research writing form and refine both individual and collective academic identities. Academic selves are (re)produced, relationally, via exchanges, collaborations and growing degrees of trust with others (Ey et al., 2020). Regardless of career stage, we are in a constant state of ‘becoming’ scholars, as we ‘shift and create new identities for ourselves, despite the seemingly hegemonic power of dominant discourses and governmental practices’ (Waitt and Gibson, 2013: 77).

Meanwhile, a focus on writing as social practice also establishes an awareness of ‘audience’. Fears and other affective experiences of the research writing process are brought back into perspective, while scholarly ideas are created and refined (Cameron et al., 2009; Heyman, 2004). For example, radical geographical (and particularly feminist) interventions (Laliberte et al., 2017; Mountz et al., 2015; Pratt, 2010; SIGJ2 Writing Collective, 2012; UKCPWG et al., 2015) reveal the social dimensions of research writing as strategically important in alleviating stress. Engaged pedagogies of research writing connect sociality with the classroom. This challenges geographers to consider how traditional social roles assigned to teacher and student around research writing can be transgressed in productive ways that benefit all participants. For Dombroski et al. (2018), the social practice of writing began with the drafting of individual texts for a co-authored book review, drawing on the experiences and insights developed through a reading group. Collective reflections on tentative textual forays catalyzed deeper insights. Social practices of writing allowed participants to discover shared discourses, experiences and emotions. Through their writing, the reading group members were able to focus ‘on the emotional aspects of [their] experiences’, decentring the emphasis on intellectual development that university processes often prioritized (Dombroski et al., 2018: 86). Furthermore, social practices of writing allowed Dombroski et al. (2018: 91) to shift their . . . attention from the “I” of self-care, personal resilience and responsibility so typical of our times, to the “we” of caring for each other in ways that enhance the wellbeing of our doctoral scholars and wider disciplinary and academic communities. The social practices of writing for the members of the reading group transgressed individualistic understandings of academia and invited participants – students and staff alike – to envisage collaborative ways of being scholars (see also Ey et al., 2020). Meanwhile, academic supervisors were able to provide the necessary care to their students as well as care for each other. This balance would have been difficult to achieve within traditional, unidirectional supervisor– student relationships. Geographers seeking to respond to the institutional politics of research writing by shifting from independence to interdependence, and from isolation to connection, should seek out ‘how care is always-already distributed between multiple human and non-human enablers’ (Dombroski et al., 2018: 91). Geographers pursuing experimental interventions around social practices of writing should work to ‘strengthen and thicken’ such care-full connections. Meanwhile, geographers should not lose sight of the fact that social connections and practices that support and nurture our scholarly writing take time to mature. 2 Reimaging the temporalities of research writing In addition to questioning static or linear understandings of research writing, an engaged pedagogical approach rejects machine-like and ‘speeded-up’ governmentalities associated with rapid, metric-driven scholarship (Berg and Seeber, 2016; Kuus, 2015; Mountz et al., 2015; Wajcman, 2015). Pedagogical approaches have reimagined the temporalities of research writing. They include encouraging open-ended explorations of diverse formats, including creative works and sustained, longer monographs. Others have broken down the ‘black box’ of the writing process into the multiple, iterative, and non-linear steps of drafting, editing, refining and polishing (Dowling et al., 2012). Interventions intensifying scholarly writing in time include reserving specific periods of time in which to challenge/commit oneself to produce a number of words, hours or projects (e.g. ‘AcWriMo’ (PhD2Published, 2018), thesis ‘boot camps’ (Mewburn, 2015) and writing retreats (Grant and Knowles, 2000; Murray and Newton, 2009)). Notwithstanding the benefits of reimagining the temporalities of research writing, engaged pedagogies demand that we bring a critical reflexivity, situating such interventions within pervasive institutional politics. For instance, while undertaking a PhD via a series of publications introduces students to ‘the particular rhythm and scheduling of writing activities’ (Dowling et al., 2012: 301), it also risks cultivating a very particular kind of (neoliberal) writerly citizen attuned to the machine-like production of paper-length outputs (Dufty- Jones, 2018). Engaged pedagogies draw attention to the tensions between mentoring PhD students towards successful careers, while cultivating critical subjects capable of critically engaging with and resisting the politics of research writing. On the one hand, more pedagogically informed and practical research training exercises are needed to better prepare candidates to survive and thrive within the speeded-up, neoliberal academy. On the other, such exercises may merely confirm the ‘rules’ of the speeded-up system. And in so doing, opportunities may be missed to challenge or resist the institutional politics of research writing that require novel strategies in the first place. A similar tension exists around ‘slow’ approaches to scholarly writing. Critical and feminist geographers have argued that slowing down research writing offers progressive and collegiate means of challenging the contemporary politics of research writing (Berg and Seeber, 2016; Hartman and Darab, 2012; Meyerhoff et al., 2011). Mountz et al. (2015: 1252) ‘slowed down’ and made time to check for militaristic and ableist language, which became ‘a calming and grounding experience with political meaning’ (Mountz et al., 2015: 1252). In a similar vein, Grant and Knowles (2000) and Knowles and Grant (2014) developed a week-long writing retreat approach to create dedicated blocks of time for research writing and ‘time out’ from the multiple and daily demands academic women face at both work and home. While productivity was an outcome, it was achieved through an ethics of care that prioritized the pleasures of writing and emphasized ‘slowness’ (Grant, 2006: 485). These and other ‘slow’ strategies aim to rethink and challenge the combative and fastpaced work rhythm that serves the metrics and efficiencies dictated by many institutions. Yet, capacities to resist speeded-up expectations of research writing are unevenly distributed and refracted by job security, ethnicity, gender, age, caring responsibilities, and so on. Likewise, ‘slow’ writing may not result from a conscious decision to resist but arise from the multiple other demands that inhibit the time needed to progress research writing. The writing of this article has taken close to 6 years – not because we wished it to mature slowly, but because pregnancy, demanding teaching and governance responsibilities, illness, family caring needs and global pandemics got in the way. Frustrations with this pace were amplified by institutional expectations of continued writing output productivity despite these circumstances. Nevertheless, with the benefit of hindsight, we admit that delays allowed for ideas and words to percolate, resulting in (we hope) a better article. While tensions can arise through efforts to reimagine the temporalities of research writing, such efforts remain worthwhile. As Power and Bergan (2019: 4332) remind us, ‘care is not always done well’ (see also Madge et al., 2009; Newstead, 2009; Tronto, 2013). Carefull approaches to teaching scholarly writing can consolidate the institutional politics of research writing. An engaged pedagogy insists that our pedagogical innovations around research writing remain alert to the risks of care. A healthy degree of critical reflexivity accounts for how neoliberalized conceptions of academic space–time inform our research writing (Castree, 2009). Engaged pedagogies challenge geographers to critically reflect on the emphases we place on different writing outputs, their temporalities, and accompanying long-term values around research writing and academic worth. 3 Reimagining the spatialities of research writing Reimagining the spatialities of research writing is our third practical strategy for an engaged pedagogical approach to research writing. Despite a wider spatial turn occurring in education studies (Comber, 2015; Fenwick et al., 2015; Gulson and Symes, 2007), space remains a nascent concept within the pedagogies of research writing, compared with social and temporal dimensions. Yet the care-full values, relationships and practices inherent to engaged pedagogies of research writing are deeply contextual and place specific: they are structured, embodied and felt through the material world (Madge et al., 2009; Noxolo, 2009). Responding to this, we propose a reimagining of the spatialities of research writing. Reimagining space invites those working in the pedagogies of research writing to become more attentive to the dynamic, affective, embodied and material roles of space and what this means for how we learn and practice research writing. It expands our spatial understandings of writing beyond ‘putting pen to paper or finger to keyboard’, to include the diffuse, mobile, precognitive and relational (Brace and Johns-Putra, 2010: 404). Learning how to do research writing should be seen ‘as something that happens through [rather than in] place’ (Saunders, 2016: 150 emphasis added). Reimagining the spatialities of research writing requires us to recognize and integrate the range of local and trans-local connections10 sustained with other people and places that variously and often unpredictably contribute to how we learn to write (DeLyser, 2010; Rogers, 2011; Saunders, 2013; 2016). As Saunders (2013: 295) explains, we can understand scholarly writing (inclusive of our texts, practices and pedagogies) as ‘a constellation, or meshwork, of social influences, encounters and tensions that emerge from our being-within-the-world’ (Saunders, 2013: 295). This approach eschews a deterministic understanding of space; space is productive, yet its effects are always unfinished and uncertain. Reimagining the spatialities of research writing in this way connects to work in higher education research on the affective and embodied dimensions of doctoral writing (Aitchison and Mowbray, 2013; Burford, 2017; Peseta, 2001; Ward, 2014). Such studies challenge the way emotions and physicality of research writing are positioned outside of the ‘master discourse’ or the ‘mind–body split’ associated with enlightenment ideals of rationality. While our bodies and emotions are unpredictable and difficult to describe and measure, they still have an enormous impact on how we learn to write. For example, drawing on Merleau-Ponty’s phenomenology, O’Connor and Petch (2012) point to the way in which the body forms a foundational space from which we develop writerly habits and dispositions (see also Barnacle and Dall’Alba, 2014; Watkins, 2017). Other researchers in higher education studies have considered how our writerly bodies take their cues from and also change the meaning of a variety of spaces (Johnson et al., 2017). Charteris et al. (2016: 41) argued that the ‘location of the academic subject’s physical body in particular spaces is semiotically coded with the value of the academic subject’s knowledge and may augment or diminish the body’s capacity to act’. Academic subjectivities are produced through and entwined with the provision of material and spatial resources (e.g. security of employment and income, office space, a parking location, and so on) and institutional spaces are key to catalyzing threshold moments in how scholars ‘become’ research writers (Charteris et al., 2016: 32). Similarly, Middleton and McKinley (2010) illuminated how complex and multiple spaces informed and were negotiated in the research and supervision of Ma˜ori doctoral students. Space and its ambivalence potential is critical when first nation people undertake doctoral research in colonial institutions. While recognizing that ‘the materiality of a space does not determine its use or pedagogic effect’ (Reh et al., 2011: 85), reimagining the spatialities of research writing through an active politics of becoming (Connelly, 2002) allows us to understand that how we learn and practice our scholarly writing is enacted relationally and spatially through nonlinguistic, unconscious, and affective meanings and practices. If we ignore our bodies and emotions, our pedagogical approaches to research writing will remain incomplete and inadequate. Being ‘comfortable’ at one’s desk (if indeed, one has a dedicated desk) is only half the trick. The other is to question ourselves in situ, mindful of others and of how spatial relations might be differently conceived. Reimagining the spatialities of writing in an engaged pedagogical fashion thus opens up opportunities to confront and reform the politics of scholarly writing. As O’Connor and Petch (2012: 96) ask, how might we ‘best cultivate environments that enrich and deepen [our own] and our students’ writing experiences’? Certain spaces on campuses might better foster productive social relationships among students, supervisors and other faculty (Dowling and Mantai, 2017). There are also considerations of campus materiality: noise, the temperature of offices, and the increasingly vexed question of openplan academic offices (Baldry and Barnes, 2012). What socio-spatial challenges to research writing practice arise from working predominantly at-home? Different spaces, occupied by different writerly bodies, generate or ease corporeal and affective discomfort that can come with learning and practicing our research writing. Engaged pedagogies of research writing draw our attention to how various geographical contexts inform and shape the quality of care (Bondi, 2003) that is offered and delivered when doing research writing. Engaged pedagogies of research writing also query what are the different rhythms of research writing enabled by and within different spaces. There may be productivity gains when taking writing off-campus and into home spaces, refracted by and woven in between domestic labour and caring responsibilities. Yet, relocation of writing off-campus may hinder the growth and renewal of peer connections, inhibiting important social practices and processes by isolating individuals. Such matters require tenured, oncampus academics to also consider how those working ‘off-campus’ secure suitable writing space, particularly those juggling precarious roles, home duties and interruptions, moving around the house and writing within daily rhythms of home and academic work. This raises another set of concerns about work/home boundaries, the permeation of working responsibilities via digital technologies into the private realms and time-spaces of the home (Gregg, 2011), and the burdens on scholars who are also carers (Klocker and Drozdzewski, 2012). An ongoing discussion is whether alternative off-campus spaces – such as those offered by the writing retreat or even the local cafe – offer practical alternatives to creating space for research writing, while managing our messy more-than-academic lives.11 Rae coordinates monthly ‘mini’ (one day) writing retreats for a group of doctoral, early- and mid-career female geographers. Like Dombroski et al. (2018), Rae uses these ‘mini’ writing retreats to create the material and metaphorical space she needs to write and to cultivate collegiate relationships that support both her own and others’ writing craft. There is no curriculum, all that is required is: one day a month; a basic timetable12; a room (of one’s own); and a nearby caf’e for breaks. The implications of engaged pedagogies of research writing therefore extend beyond how we might change our research writing practices and the politics therein. They have the potential to affect every aspect of our professional lives. Engaged pedagogies of research writing create and reinforce opportunities and practices that work to replace dominant subjectivities framed around homo academicus (the ‘academic person’) with a collectivized identity of discipuli curans (i.e. ‘caring scholars’). The singular is replaced with the plural and what it means to be a member of faculty shifts: from being an individual within a hierarchical elite to becoming part of a scholarly collective committed to continual curation and discovery.

V COVID Coda We concluded major revisions to this article in April and May 2020, just as the SARS-COV-2 virus was declared a global emergency, and worked again on minor revisions a year later, amidst the ongoing pandemic. Under initial lockdown measures, frantic efforts were underway to deliver teaching online, while a year on, rapidly deteriorating financial scenario at universities cast new doubts on the resilience of the entire sector. We have ceased to write from our campus offices and our scholarly lives have become confined almost entirely to our homes. These are the prosaic spaces where we and our family, friends and colleagues adjust to isolation and deal with the fear and reality of loss—loss of life, employment and well-being. Our homes have become classrooms for our students and our children, our bedrooms and dining tables backdrop to countless zoom meetings and online synchronous classes. Our lives and the lives of those we care with have been dramatically disrupted and are unlikely to be the same again, socially, economically, culturally and politically. Within this context, our years of thinking and writing about the politics of research writing further crystallized. Previous global (economic) crises have shown that neoliberalized governmentalities prioritizing metrics, competition and individuals are hopelessly inadequate but distressingly enduring when it comes to matters of safety and survival, life and death (Peck, 2010; Smith, 2015). The COVID-19 pandemic has once again brought to the fore the urgent need to think and act collectively. In these circumstances, the cult of the individual cannot prevail; when it has, it has done so at the cost of hundreds of thousands of lives. From this vantage point, we see with renewed clarity how elements of contemporary higher education and the politics of research writing have weakened our institutions. Vice- Chancellors have implored their academic staff to ‘look after’ and ‘support’ each other, blind to the fact that their investment in academic cultures of competition combined with evergrowing levels of precariousness in many academic workplaces has depleted the reserves of altruism and collegiality needed to sustain their institutions. While a lucky few have welcomed an ‘unplanned sabbatical’ (in the words of one quip made by a colleague with few teaching responsibilities), horror stories abound of exploitation and unsustainable stress, as universities reply upon academics to over-work in an emergency response, all-the-while fearing impending redundancy due to ever-evolving funding crises (Marshman and Larkins, 2020). Having to rewrite curriculum and shift modes of course delivery on the fly, academics (often women with children, on short-term rather than continuing contracts) are abandoning research and writing projects to answer late-night emails or calls from students in crisis, working weekends to meet ever-shifting daily and weekly deadlines. One colleague and close friend woke at 3 a.m. to record lectures in her kitchen pantry before homeschooling her children during the day. The COVID-19 pandemic has demonstrated how institutional politics, neoliberal management and metric dogma pose a serious threat to not only the wellbeing of individuals but also the resilience of communities of scholars, and the long-term sustainability of higher education. Just as the wider world might never be quite the same because of the COVID-19 pandemic, we see important opportunities at this moment to change how our scholarly writing is governed and valued. The COVID-19 pandemic may well intensify the calculative technologies that seek to guide how we perform and produce our writing. University jobs are being cut because of pandemic-related financial crises, institutions are making these decisions based upon the very same metrics around research writing that critics have challenged for many years (Marshman and Larkins, 2020). Yet, it is also true that these new circumstances – while undoubtedly constrained – also offer opportunities for contestation and resistance. Our writerly tasks are increasingly confined, codified and surveilled – literally in the sense of having to write from isolation, and with Zoom, Webex and other online teaching and networking platforms susceptible to management monitoring. Yet, as subjects – as selves – we constantly ‘become’ who we are ‘through power relations that both impinge upon us and activate us’ (Waitt and Gibson, 2013: 77, emphasis in original). With numbers of grant applications, publication submissions and quality peer review likely to fall off the cliff as academic workloads become completely unmanageable, the crisis may well force an overdue recalibration of management priorities. Rather than view present struggles to write as ‘individual failures’, now is the time for prosecuting collective narratives and acts of care beyond the ‘win-at-all-costs’ mentality.

Professor Akilagpa Sawyerr, during his tenure as Secretary-General of the Association of African Universities 2003–2008, described the history and status of a steeply inclined, uphill struggle to strengthen institutional research capacity across sub-Saharan Africa from the time of Independence (1960s). He envisioned ‘… sustained indigenous generation of world-class research results and new knowledge that help our understanding of African conditions and contribute to the advancement of its people’. (Sawyerr, 2004, 215). Research would be conducted by African researchers working primarily at African institutions, turning out first-rate knowledge on locally relevant issues, contributing as well to global knowledge. His analysis, and that of many others then and since (Garfield, 1983; Salomon et al., 1994; Ajayi et al., 1996; Beintema et al., 1998; Task Force on Higher Education and Society, 2000; Crossman, 2004; Oyewolea, 2009) indicated major changes that would be needed in African societies (macro conditions) and within African universities (especially public ones, still the hubs of research in their countries), and by international collaborators if anything like this vision were to be realized. Another grand man of Africa and academia, Professor Thandika Mkandawire (former Director of the Council for the Development of Social Science Research in Africa (CODESRIA) and Olof Palme Professor for Peace with the Institute for Future Studies in Stockholm) injected into debates ongoing in 1995 his historical perspective, describing three generations of ‘indigenous’ African researchers in post-independence Africa. His observations, still valuable today, seem especially appropriate, given that this article will present research capacity strengthening contributions of three generations from a group at Makerere University’s Faculty of Medicine (Department of Pharmacology and Therapeutics, now in the College of Health Sciences) and its collaborators. Ugandan academy-based researchers including three authors, one from each generation — Professors Jasper Ogwal- Okeng (JO-O), Paul Waako (PW), and Celestino Obua (CO) — have pursued an evolutionary course in doctoral education, research, and institutional capacity strengthening from the 1990s, yielding achievements and lessons presented below. To set the scene, we go back to earlier periods of post-colonial African history. Mkandawire’s first generation, African students of largely ‘cold-war induced airlifts’ to American universities, studied at some of the best institutions in the United States and in Europe (Mkandawire, 1995, 75), most shortly after independence in their respective countries. Many returned home after their studies, soon to become the senior African scholars and mentors. Compared with what was to come, the first group returned to relatively functional and expanding institutions where donors and governments invested considerable resources, although it was already possible to detect the ‘seeds of authoritarian rule’ (Mkandawire, 1995, 76). His second generation often completed undergraduate degrees at their national universities, pursued post-graduate studies abroad, but few came home — giving way to ‘brain drain’. From the mid-1970s — and into the 1990s, ‘academia and the state were virtually at each other’s throats (literally in the case of the throats of the former)’ (Mkandawire, 1995, 77). Further complications included a global recession, International Monetary Fund (IMF) and World Bank ‘structural adjustment policies’, inflation rates and devaluations of currencies. These pressures reduced local salaries, and in many universities, prohibited new hiring to the point that few of the growing diaspora of African scholars could be recruited home (if political repression did not otherwise keep them way). It is not difficult, then, to imagine how many African universities became dilapidated images of their former selves. That is, they struggled to keep up teaching in deteriorating physical facilities too small for growing numbers of undergraduate students — in the absence of modern libraries, information communication technology (ICT) to connect to scholars worldwide, laboratory facilities or equipment, with cuts in water and electricity. A paucity of highly trained scholars or scientists made it impossible in many places to ramp up high-quality post-graduate programmes, or even to mentor effectively students with aptitude and interest in research. Most of Mkandawire’s third generation trained in their home universities, first as undergraduates, then for advanced studies; there were few resources from African governments or international sources.1 Especially those who began post-graduate studies late in the 1980s trained under extraordinarily difficult circumstances, having come of age under repressive regimes—as products of independent Africa, but with limited academic freedom and ‘scandalous material conditions’ (Mkandawire, 1995, 79). By 1995 this group had begun ‘assuming the reins of power’ in their universities, as mid-level and senior members of the academic hierarchy (Mkandawire, 1995, 79). How can universities in Africa and elsewhere strengthen their capacity for research? The major contribution of this article is a case study at Makerere University from which we draw lessons. We set the case study in context, noting key points about the evolution of research as an element of national development and development assistance. We then turn to universities and their research contributions, with special attention to the role of doctoral studies worldwide and in Africa. Next we introduce Makerere University. This leads to the institutional research capacity strengthening collaboration between Uganda’s Makerere University and Sweden, the largest source of external funding of Makerere University, 2000–2012. The case study illustrates how a department within Makerere’s Faculty of Medicine contributes to transforming the post-graduate educational process with formation of a productive research team among faculty and with students to advance a research agenda targeting strategic priorities of national significance. We conclude with lessons and recommendations about capacity building for African universities and their collaborating partners. Context: Development Strategies, Universities, Doctoral Studies, Research, and Institutional Research Capacity Research as a planned component of development Universities in Africa, as elsewhere in the world, are the primary means for educating leaders and workforces to be able to take on the challenges of development. In all societies, research has been a way to meet strategic national goals, albeit different ones in countries early to industrialize, and those doing so today. For countries early to industrialize, research, largely in universities, has been a component of development that evolved from industrially driven growth (1870s–1930s). European and North American nations invested in research to advance sectors seen as strategic, including agriculture for food self-sufficiency, energy, transportation, communications for domestic and foreign policy goals, including military might. Economic competitiveness came to characterize the common goal of the more affluent countries, such as those in the Organization for Economic Cooperation and Development (OECD). Once independent, countries in sub-Saharan Africa remained largely exporters of natural resources, not industrial commodities. In 2003 an African Ministerial Conference on Science and Technology stressed the urgency of building the continent’s science and technology capacities to eradicate poverty, fight diseases, stem environmental degradation, and improve economic competitiveness (Funke and Nsouli, 2003). The African Union recommended government investment in research and development (R&D) of 1% of GDP (AU — NEPAD, 2005; AU — NEPAD, 2010); OECD countries aspired to 3% (OECD, 1971). Research (or science and technology assistance) as development assistance Discussion of research for developing countries moved into the domain of development assistance. The World Bank packaged most aid for science and technology for larger countries (China, Indonesia, South Korea, then Brazil) bypassing sub-Saharan Africa (Muskin, 1972; Thulstrup, 1999). Another potential source of support for knowledge generation, the world’s large science agencies, including medical research councils and institutes of health, concentrated on their primary mandate, advancing science for application at home. This was true even when they conducted or funded research in developing countries (Freeman and Miller, 2001). Research agendas driven by international network members in control of most of the funds meant topics were not often well suited to the problems of highest priority for development of African nations. Power imbalance in research collaborations thus reduced their development potential (Gaillard, 1994). Although most developed countries largely avoided supporting research rooted in less developed ones, Canada and Sweden proved exceptions. They created the only two development organizations dedicated to research. After World War II Sweden recognized that ‘knowledge’ played a central role in its own dramatic development into an industrial power, then also promoted research to enhance economic growth abroad (Larsson, 1995; Wijkman, 1995). In 1975 the Swedish parliament founded a special body for research cooperation within the Swedish International Development Agency (SIDA), and later created a separate agency, the Swedish Agency for Research Cooperation with developing countries (SAREC). Sweden became one of the world’s foremost proponents of institutional research capacity development for developing countries — a term intended to distinguish support for institutions to create modern research environments from that of support for individual scientists or projects. Most other development agency support labelled as ‘capacity development’ principally meant aid for managing programme operations, not for generating knowledge (Freeman and Miller, 2001). Global recession, coupled with structural adjustment policies of the International Monetary Fund (IMF) and World Bank, meant most international funding for education poured into primary schooling. Governments seeking World Bank loans were instructed not to invest in higher education — as debtor nations they were to restructure economies away from social investments (Bloom et al., 2006). Thus, investment in universities waned until 1999 or later and research capacity suffered without money for these primary engines for advancing knowledge (Bloom et al., 2011). Underfunding of higher education in Africa by national and international leaders seemed ‘especially paradoxical’ as it coincided with the ‘… very moment of the rise of the knowledge society and its spread throughout the globe’ (Sawyer 2004, 236). When the IMF and the World Bank moved away from structural adjustment, they introduced in 1999 a strategy for poverty reduction based on comprehensive, country-based poverty reduction plans (IMF, 2013). Universities, research and doctoral studies worldwide Today universities worldwide focus on teaching, service, and research, with research as the major source of prestige. Research output and its benefits for economic competitiveness underlie systems for comparing, or ranking of universities within and across nations. Indicators used in global ranking systems suit the development of, and ongoing competition for, economic power: publications in high impact journals and patents. These systems of comparison (e.g., Thomson-Reuters,http://www.timeshighereducation.co.uk/world-university-rankings/2012-13/world-ranking, QSWorldUniversityRankings http://www.topuniversities.com/qs-world-universityrankings, and Webometric Ranking of World Universities, http://www.webometrics. info/en/world) do not suggest a developmental path for universities in settings where research resources have been few, and patents not necessarily so useful a measure of research well suited to poverty alleviation nor for human as well as economic development. These rankings do not measure growth in quantity or quality of activities or of outputs for which universities in early stages of building research capacity might be usefully assessed to show positive change over time. For example, conference presentations and publication nationally or regionally often precede success in publication internationally, especially in higher impact journals. The roots of universities in 12th century Europe and their roles in research and development have been described elsewhere.2 Doctoral education evolved from the Doctor of Philosophy, based on research qualifications introduced by von Humboldt in Germany in the early 19th century (Pearson, 2005). Major political, economic, and social changes led to education taking place in research laboratories, through seminars and with senior professors as tutors and heads of departments (Gellert, 1993). Then American institutions in the 19th and early 20th centuries modified doctoral education, combining doctoral coursework with research, organized in graduate schools with undergraduate education in the same departments (Gumport, 1993). The Ph.D. degree spread from Germany and the United States to other countries, often overtaking already existing research degrees, for example, ScD, DA, DLitt, DPhil, MA by thesis, MLitt, by research in the United Kingdom (Jörgensen, 2012) and processes — by publication, taught doctoral programmes, and ‘professional’ doctorates (Boughey and McKenna, 2013). Europe has recently changed doctoral education towards a model similar to those at US universities, based on more teaching and with progression of studies from bachelor through Masters programmes, then the Ph.D. degree. African universities and doctoral programmes Most African universities have followed similar developmental paths, incorporating varied academic traditions from the UK, France, Belgium, Portugal and others. After independence the number of graduates was very low; French-speaking African institutions produced only four doctoral graduates of agriculture and English-speaking Africa produced 150 in the period 1952–1963 (Eisemon, 1982). Low enrolment in higher education compared with other parts of the world persisted; less than 3% in Africa (Teferra and Altbach, 2004). Recently African universities have been moving from elite to mass education (Perkin, 2011). Even so, total yearly expenditure for higher education in all of Africa is lower than expenditures of a single one of the richest universities in the United States (Teferra and Altbach, 2003). Some countries, including Uganda and Tanzania, reduced public support for universities while increasing enrolment by admitting more private (self-paying) students (Weber, 2005). Increase in locally trained Ph.D.s grew from an unacceptably low base (Sawyerr, 2004). A study of six universities in sub-Saharan Africa (Cameroon, Nigeria, Kenya, Benin, Senegal, Rwanda) found a 300% increase in doctoral students (from 373 to 1,454) from 2005 to 2009 (van’t Land, 2011). Makerere University History Makerere, prior to becoming the national university of Uganda, had a more promising early history than many public research institutions in Africa (Musisi and Muwanga, 2003; Sicherman, 2005). From its start as a technical college in 1922 it was a strong institution in the region and became part of the University of East Africa in 1963. From 1970, when the University of East Africa split into three independent universities, Makerere offered undergraduate and post-graduate courses and awarded its own degrees (Sicherman, 2008). Political stability and the hopes it created brought new research initiatives, the Medical School research programme at Mulago Hospital among them (Sawyerr, 2004, 238). From 1970 to 1986, as it evolved from a British-dominated, regional institution, Makerere changed its curriculum and teaching methods, introducing locally relevant material and interdisciplinary studies. During those same years, the university suffered from the repressive rule of Idi Amin, whose regime particularly targeted Makerere, abducting and killing students and staff. Brain drain ensued as academics fled. Makerere’s financial resources from both public and external sources declined dramatically in the 1970s and 1980s — just as the university was under pressure to expand enrolment. Institutional advances of the 1950s and 1960s declined and recovery began slowly in the 1980s (Doro, 2007). When Yoweri Museveni came to power in 1986, Makerere expanded its admissions and curricula. Staff development programmes began to reverse the brain drain, and to ‘Africanize’ the staff. But the economic situation did not improve quickly and salaries remained low — considerably lower than for institutions in neighbouring countries (Doro, 2007). Makerere University and higher education in Uganda today Five public and 29 private universities operate in Uganda today. Of these Makerere University remains by far the largest with 34,694 registered students in 2011 (Makerere University, 2012, 14). Of these, postgraduate students constitute about 5% of total enrolment; over half of these study humanities and social science (Makerere University, 2013, 26). Uganda has currently one researcher per thousand members of the country’s workforce, compared with more than five in the developed world (Bailey et al., 2011). As of 2010, research conducted in Ugandan public universities was largely internationally funded — by Sida, Carnegie Corporation of New York, the UK’s Department for International Development (DFID), Norway’s NORAD, and USAID in the United States (National Council for Higher Education (NCHE), 2011). Makerere University ranks ninth in Africa with regard to scientific output, according to the Schimago Institutions Rankings (SIR) in 2012 (http://www .scimagoir.com; see also Kiwawulo, 2013). Growth of post-graduate education at Makerere from 2000 Historically most Ph.D. holders at Makerere University studied in Europe or North America. Some European programmes, particularly the ‘sandwich modality’ already in use in the early 1990s by SAREC (Bhagavan, 1992) and common in other Nordic countries, combined coursework, use of libraries and laboratories with mentoring in the European country — with fieldwork in the students’ home countries. A very small number pursued UK-style Ph.D.s by research at Makerere, but completion was uncertain, with many variables beyond the control of the students, and even of supervisors. Thus few initiated doctoral studies at Makerere prior to 2000, and even fewer completed them; rates were higher for degrees granted abroad. Research funding, mostly from abroad, and historically mostly allocated project by project, meant that better financed partners used their resources to leverage choice of topics and roles (Gaillard, 1994; Nankinga et al., 2011; Okui et al., 2011). A major turning point came in 2000, both for research funding, and for the nature of major support. Senior academicians reported that Makerere started to move away from its historically ‘donor-driven’ research agenda when deans, directors and the Vice-Chancellor discussed among themselves, in 1999, ideas for a programme of support offered by Sida (Freeman et al., 2010). Research Capacity Strengthening: Case Study of Collaboration Uganda- Sweden3 Collaboration Sweden-Uganda: Makerere university-Swedish universities Sweden entered into bilateral cooperation with Uganda in 1999. SAREC4 initiated an analysis of conditions for research cooperation. Commissioned by Sida/SAREC, the Ugandan National Council for Science and Technology (UNCST) participated in analysis of the research landscape that identified Makerere as the institution performing the most research and the only one with the right to grant Ph.D. degrees (see Figure 1 for Sida approach). To learn about research-related interest in the units, in 1999 Sida/SAREC consulted all Faculties and Directorates and many individual departments in visits at Makerere University. Sida/SAREC and the Makarere leadership invited all Deans and Directors at Makerere to clarify a focus for the cooperation. At a meeting participants came to a consensus to support research capacity strengthening emphasizing local Ph.D. training. Makerere would grant the Ph.D. degree, students would study using the ‘sandwich’ modality.5 Consensus also led to selection of a theme, ‘Lake Victoria and other water bodies’, to encourage collaborative research. They adopted a slogan of ‘Supporting the supervisor to supervise’ to concentrate on improving supervisory competence of Ph.D.-level researchers at Makarere. Sida/SAREC focused its support on relatively strong faculties that had Ph.D. degree programmes. It awarded planning grants to the Faculties of Agriculture, Medicine, Social Sciences, and Technology for each to prepare a proposal for Sida/ SAREC support. Considering the central importance of Information and Communications Technology (ICT) for the functioning of a modern research university, Sida/SAREC supported Makerere to develop an ICT master plan to clarify need and negotiate donations more effectively from any external organization. The investment in hardware for university-wide ICT infrastructure transformed the functioning of the university (Tusubira and Mulira, 2004) including major modernization of the library (Greenburg and Versluis, 2005) and Uganda-wide access, for the first time, to more than 20,000 full text electronic journals in all disciplines. Methods for assessing institutional research capacity gains Two of us from an evaluation team commissioned by Sida in 2009 (Phyllis Freeman (PF) and Eva Johansson (EJ); a third evaluator undertook the financial audit) sought evidence of contributions of the Sida-funded research collaboration to Makerere University’s strength as a research university (2000–2008). The evaluators searched for an impartial approach, as after monitoring progress of the groups since 2004 (under contract to Sida), Freeman and Johansson were familiar with all of the participating research faculty and students. We found a promising method, developed at the London School of Hygiene and Tropical Medicine (Kuruvilla et al., 2007). The data collection instrument called for factual and verifiable information and offered an easily manageable way to display and compare results. This would allow funders, participants, and observers alike to review the data and compare results. As the purpose of its designers was comparative evaluation of research projects, these evaluators added categories to the original template for assessing change in institutional research capacity. Freeman and Johansson completed 12 such assessments of research groups in 2009 (from the Faculties of Agriculture, Technology, and Social sciences as well as Medicine and Public Health) and reported results to Sida, including to the senior Sida staff member originally in charge of the programme (also an author here, Hannah Akuffo), and to Makerere University. The comparison alerted evaluators that it was the Department of Pharmacology and Therapeutics (hereinafter ‘Department’) that could offer a most encouraging example from which others might learn. Sida distributed the draft evaluation report and arranged for an open discussion of the results at Makerere University, attended by Johansson, before final corrections and publication of the report in 2010.6 Sida also published the complete version of all data collected for comparison in full detail, including names, amounts, and duration for all grants and contracts during the study period, all publications, staff hired, promoted, laboratory enhancements, etc.7 In 2009, Waako (PW) also pointed to his predecessor Ogwal-Okeng (JO-O), then founding dean of a new medical school upcountry in Uganda, in Gulu, as the best source for a historical view of how the Pharmacology research group evolved. PF interviewed JO-O and summarized the developmental history of the group in 2009 (as Appendix Q to Report Annexes cited in Note 7, 736–741) to facilitate understanding of how the pharmacology group evolved. In preparation of this article, PF investigated whether the pharmacology research group had continued to advance in ways that could usefully encourage others. She gathered factual and verifiable updates to the 2009 template (see Box 1 for a summary) and further insights from PW and JO-O (included in the section below entitled Genesis of a Flourishing Research Group). They suggested including the first Ph.D. graduate from the programme of collaboration with Swedish universities (CO) as an author as the best way to add a new perspective. PF then recorded the updated perspectives from JO-O, PW, and CO for this article — representing three generations in the genesis of the group — and its importance for the Faculty of Medicine, the College of Health Sciences, and Makerere. Research Capacity Growth in the Department of Pharmacology and Therapeutics, 2000–2012 Outcomes In 2000, just as the collaborative Ph.D. programme opened, the Department’s research group responded to the introduction of new drugs for malaria and HIV/ AIDS for use in large numbers of patients across Uganda. They established that increasing the dose of sulphadoxine/pyrimethamine (SP) improved efficacy and maintained safety of chloroquine-SP treatment for malaria. In addition, a serious deficiency in information for clinicians led them to establish the Drug Information Service, free to clinical providers. Since 2006 they have been contributing to policy formation as members of key national policy formulation bodies. They collaborate with colleagues at Makerere and with international colleagues. The findings directly influenced WHO/UNICEF policy to improve treatment not only of malaria, but also of pneumonia, an equally large cause of death that has long been neglected. The group is expanding services of the Information Service to include poisons. Other notable achievements in institutional terms are the accomplishments of the first faculty and early cohorts of Ph.D. graduates from the collaboration. Many have taken leadership positions in Makerere and the Ministry of Health: CO is Deputy Principal of the College of Health Sciences, Mohammad Ntale is Chair of Chemistry in Makerere, Jesica Nsungwa is a Commissioner of Child Health in the Ministry of Health, and, with PW soon to become the first Dean of Busitema Medical School in eastern Uganda, Jackson Mukonzo is taking over the Department Chairmanship in Makerere. JO-O, even while starting the Gulu medical school in northern Uganda9 from scratch, won a Challenge Grant from the Gates Foundation (for original field research). The capacity gains at the national university are thus spilling over to benefit public universities beyond the capital city. Genesis of a Flourishing Research Group in the Department of Pharmacology and Therapeutics First we explain the process through which JO-O, the first generation among Ugandan authors here, and the originator of this programme became a researcher and Ph.D. holder himself (updated from Freeman et al., 2010, Annex Q, 736–741); then we describe how he and PW of the second generation built the growing group whose work is summarized in the Institutional Research Capacity Gains Summary, including the perspective of our third generation, CO. Observations of a general practitioner led to research topics In 1998, JO-O completed a five-year journey resulting in a Ph.D. by research. At the start, the only option was to organize an individual course of study by research—there was no Ph.D. programme. With the expulsion of Asians from Uganda by Idi Amin in 1971 and departures of most expatriate faculty, researchers had become scarce. JO-O’s background included a first degree as medical doctor (1978), followed by two years at the Mulago Referral Hospital (associated with the Faculty of Medicine) in charge of Casualty and later the Tuberculosis and Chest Ward. From 1980 to 1987 the Ministry of Health posted him upcountry to practice general medicine in a 100-bed district hospital in his home region of Lira. He became both the Medical Superintendent of the hospital and the District Medical Officer — responsible for planning and oversight of all public health programmes for all of Lira’s residents. JO-O returned to Makerere in 1987 to pursue an academic career in Pharmacology for public health, first as a Teaching Assistant/Graduate Fellow, followed by completion of a Master’s degree in Pharmacology (MACP) in 1989. His appointment as lecturer in Pharmacology came in 1990. He was the youngest of four lecturers in the department; two others worked on contract (over 60 years), and one went to open a Pharmacy Department in Makerere. His research interests emanated from observations as a physician. While serving in obstetrics in Lira he noticed that many women in labour held combinations of herbs in their hands or chewed them to hasten labour — a tradition passed down by their aunties. JO-O wanted to know if the herbs had chemical properties to hasten labour — this sort of curiosity led to studies for both his Master’s and Ph.D. The herbs about which he wrote his MACP thesis were those that patients took to ward off intestinal worms. As there was little focus nationally or internationally on intestinal worms at the time, when he turned to designing his Ph.D. studies, he also turned to research on malaria—a field that drew far more attention from researchers with whom JO-O could learn and collaborate. It remains one of the most significant causes of childhood illness and death in Uganda, and a world-class killer across Africa and Asia. From informal mentorship to Ph.D. registration and official supervision As the Ph.D. by research model offered no official supervision until much later in the process, JO-O pursued informal mentorship from three sources. For field studies of plants used against malaria, he collaborated with a professor of botany; then he sought assistance with plant chemistry, then in pharmacology. The relationship with each informal mentor developed serendipitously—including occasional exchange of contributions to each other’s work. The first official step towards the Ph.D. came when JO-O felt prepared to present his concept to the Faculty Board, consisting of senior academic staff, including a few Ph.D.s in the Makerere Faculty of Medicine. Then the university-wide Higher Degrees Committee issued a formal decision, clearing the way for him to register for the Ph.D. Only then did the Academic Registrar allocate two official Ph.D. supervisors — and awarded JO-O a payment of three million Ugandan shillings (today about US$1,170) to carry out his doctoral research. A Ph.D. by research meant learning research methods on his own, or if he could determine appropriate courses (abroad) and pay the costs, he could study with a group. From registration to faltering — then to completion For analysis of the chemical composition of plants used in Uganda to treat malaria— to determine if indeed they possessed anti-malarial properties — the World Health Organization (WHO) offered access to a good malaria laboratory in Madagascar. He found a lab so well-equipped and efficient that he had all samples run in three weeks, and returned home to write his dissertation. Despite the efficiency of the laboratory analysis, the overall process proved so arduous that JO-O nearly gave up. The Academic Registrar chased him down when JO-O had reported no progress for some time — and convinced him to continue. From Ph.D. to building a department and research group Upon completion of his Ph.D. in 1998, JO-O received a US$36,000 grant from Rockefeller Foundation (African Career Awards Program) to set up a laboratory — the one that made it possible for PW to conduct research for his Master’s degree (1995–1997). When JO-O also became Chair/Head of Pharmacology in 1998, his first priority was to develop research and teaching capacity. With little support for staff development, the only way one could advance younger colleagues would be through collaboration with institutions in the western world or the newly democratic South Africa. PW was the first entrant of the next generation. As PW recounted, entry into academia was stark: One morning in January 1995 I came from Iganga District Hospital to Kampala and walked to the Department of Pharmacology and Therapeutics. It had the fewest teachers during my training—three. I was welcomed with a silence that characterized the Department for quite some time — not only then but during the years to come when I was a Master’s student and Lecturer as the World Bank ban on hiring did not end until 1999. In the corridors I did not meet anybody until reaching the office of the Head of Department who welcomed me warmly and asked what he could do for me. I said I wanted to join the Department as a teaching assistant. Professor’s face beamed and he replied ‘so you would like to pass through the Department of Pharmacology’. His statement has continued to mean many things; I later learned that many doctors had been sent overseas through the department and never returned to serve there. The more positive interpretation is that many pharmacologists have risen through the ranks and hold positions of responsibility in the country. JO-O initiated a first international collaboration to advance his capacity strengthening goals with a colleague from University of Cape Town whom he met at a conference in Kampala. That Professor agreed to take PW for Ph.D. training. Supported by the Makerere staff development fund (after a year-long debate during which the committee twice rejected the Department’s application), PW departed for Cape Town. PW joined a vibrant department with 15 Ph.D. students, becoming one of eight Ph.D. candidates whom his supervisor guided. Even though that supervisor spent one of those years on sabbatical in Geneva, the eight Ph.D. students never lost focus. Regular research meetings fostered student-to-student learning. As PW recounts: ‘This was an eye-opener to me about how Ph.D. training can be conducted in a good research setting’. While studying in South Africa, PW returned to Uganda twice a year for one month. PW completed his Cape Town University degree in three years and returned to Kampala in 2003 — adding an international ingredient to the formation of the Makerere Pharmacology research team. Having been appointed Lecturer in 2000 PW assumed the Department chairmanship in 2004 when Uganda tapped JO-O to open another public School of Medicine in Gulu, a troubled region in northern Uganda. JO-O felt he was leaving the Makerere Department in very good hands and reserved 25% of his time to return to supervise Masters and Ph.D. students. New funds promote building of a cohesive research group Two international groups announced competitions for research funds: Sida/SAREC in 2000 and the Inter-University Council for East Africa (IUCEA)10 in 2003. The Sida/SAREC grant enabled the Department to collaborate with the Clinical Pharmacology Unit in Sweden’s Karolinska Institute (KI), based in Huddinge Hospital. In collaboration with researchers in Kenyatta University in Kenya, this group also won US$50,000 (renewed in 2005) for Research training in Pharmacology. As both grants involved research training JO-O recruited five Ph.D. students (four Sida/SAREC, one IUCEA). Sida’s call for proposals emphasized institutional capacity building. From the start of the collaboration with Sweden, JO-O chose to build capacity in three areas: 1. Capacity for supervision of Ph.D. students: As this would be a new activity for JO-O, his Swedish counterpart would co-supervise the doctoral research. 2. Human resources for clinical pharmacology: Makerere collaborators identified University staff with Master’s degrees to enrol as Ph.D. students to add highly educated individuals to the workforce. The first group brought backgrounds in chemistry, public health, pharmaco-epidemolgy, and medicine. Each doctoral student worked with two MSc candidates — the talent pool from which later Ph.D. students would be recruited. 3. Capacity in the Makerere laboratory: equipment and personnel to analyse levels of medicines in blood and other tissues. JO-O proposed to use Sida/SAREC funds to initiate Clinical Pharmacology in Uganda, establishing the country’s first drug analytical laboratory in the Faculty of Medicine, supporting tuition and fees, research costs (field work and consumables), travel for periods of work in Sweden for Ph.D. students, and support to the Swedish supervisors.11 He also proposed to supply research material and space for Master’s students — who would pay their own tuition and fees. This strategy worked, and eight Master’s students completed degrees this way in the initial programme — then became the new prospects for Ph.D. studies; each year two have registered. The Master’s students now make up a good portion of more than 20 researchers who, since 2007, have gathered weekly for research seminars — whether senior researchers are present or not. Today researchers from elsewhere in the university petition to present to this seminar, having heard what a good option it is for supportive critique and guidance. Supervision Although JO-O‘s own Ph.D. experience worked quite well, he wanted a different model for his Ph.D. students. An important experience for JO-O came with a Fulbright scholarship to the United States in 2002, at the University of Illinois, where he watched senior researchers who supervised informally — as well as formally (as in Uganda); American supervisors often sat by their students in the lab, discussing the student’s work in depth and guiding Ph.D. candidates on writing manuscripts. In the Makerere—KI collaboration, KI supervisors visit Uganda twice each year. The visit routinely begins with a meeting of all supervisors with all Ph.D. students; they discuss what work has been done during the previous few months and issues of common interest or concern. The environment is one of free-flowing discussion. Next, supervisors (usually three from the two institutions) sit with each student individually, reviewing the scientific merit of the work in progress—and providing guidance on next steps. The objective for joining all supervisors with each student is to eliminate a serious challenge in other departments where a student may work separately with each supervisor—and receive conflicting guidance. In a third type of meeting, the specialist KI visitor sits with the student to engage around the scientific questions in greater depth. JO-O found himself envious of the opportunities of this generation of students in the collaboration with Sweden — so different from the British tradition of hierarchical relations. He finds the new method instills a higher level of confidence in students—especially as a cardinal premise of the supervision sessions is that there is no ‘wrong’ answer during brainstorming discussions — to encourage creativity before critical analysis. As one of the first cohort of five Ph.D. students in the Department, CO observed that from collaborative supervision: … there have been institutional cultural issues that we students needed to accommodate, for example, strict keeping of deadlines at KI, or where in Uganda one cannot address the supervisor by the first name while in Sweden this was the norm. The external supervisor equally needs to appreciate these differences for the mentorship to yield fruitful results — as Ugandan students are often more mature and rather independent thinking. It has been a richly rewarding experience, and to a large extent I owe any success after the Ph.D. program to these experiences, the capacity to balance different personalities and interests of the supervisors across the continental divide while maintaining a semblance of my own. Setting the group’s research agenda At Makerere, as elsewhere, there exists a long history of collaborators from the North coming with projects already formulated, looking to engage on pre-set agendas — possibly of relatively small interest to Ugandan partners —even if well-funded, thus potentially irresistible. For the Pharmacology Department, senior investigators set topics, blending opportunities for support with Ugandan priorities, based on their years of experience as clinicians as well as researchers. Students agree to work in these designated areas, having learned from their predecessors about the Department’s favourable environment for Ph.D. study. Looking back and assessing while moving on CO assesses his Ph.D. experience and the advantages it has offered from his current position within the senior ranks of leadership of the College of Health Sciences: After my Ph.D. training I have been able to contribute to the departmental research team, supervise, and win competitive grants. It is important to be a team player; any research becomes complete when members complement each other’s efforts. I attribute my rise into administration as directly related to the Ph.D. training. I gained confidence to participate in grant writing to support further research. I have been able to publish regularly and appreciate peer-reviewed journals, especially those with fairly high impact factors. I have become a regular reviewer for international journals as well as for international grant applications. After two years on the Research and Ethics Committee for Medicine I joined the Editorial Board of the African Health Sciences journal. After learning the (KI-MU) collaboration as a student, I joined the world of institutional diplomacy, full of negotiations and brokering collaborations and partnerships with regional and international institutions- as an administrator. I now provide guidance on the KI-MU collaboration for new students and sometimes for senior researchers too (many of whom needed to better understand the KI-MU collaboration and how higher education operates in Uganda in relation to international organizations). The mentorship by dedicated individuals from the KI-MU collaboration, made possible by the Sida support, is one of the best examples of a capacity building partnership that potentially can be self-sustaining.

Conclusion: Lessons and Recommendations As of 2013, the Swedish-Uganda research cooperation agreement with Makerere University is extending benefits to other parts of Uganda, training Ph.D.s who will teach, supervise, and conduct research in other universities; expanding access to electronic journals nationwide; sharing laboratory facilities and expertise as senior Makerere faculty take up leadership roles in other public universities.12 At Makerere, Sida broadened its focus to add support for modernizing administrative and financial rules and procedures to support world-class research; this reform process is also needed to strengthen institutional research capacity. What are the main lessons and recommendations from this experience? We emphasize four crucial ones. Start-up The most important lesson, according to JO-O, is the universality of the message: he thought of himself and his colleagues as ‘ordinary Ugandans’, not somehow special or specially prepared for the work they were about to undertake. Thus, other groups can similarly build research capacity: initiate research and research degree programmes, sharing priorities and collaborating closely, and do so by seeking international partnerships that suit the host university’s goals. Recruiting Ph.D. students Watch for research interests and talent in undergraduates, encourage them to pursue Master’s degrees within a group that shares research priorities and strategies, then invite those most capable in research to pursue Ph.D.s — especially if they demonstrate commitment to contribute to the home institution or other important bodies within their countries throughout their careers. Supervision, teamwork, and research collaborations The supervision model detailed in the case study above and the weekly departmental research meetings encourage openness, creativity aided by brainstorming, as well as rigour; these have contributed to immensely productive, long-term relationships among Ugandan and Swedish scholars who publish together. These strengths also helped to develop additional collaborations, in Uganda and internationally, ones that bring fresh perspectives and funding — with exposure for younger scholars to research as conducted in other countries. Choosing collaborators requires the host institution to anticipate the willingness of research partners, especially international ones, to appreciate and respect local knowledge and priorities for choosing research topics, the importance of disseminating findings in forms appropriate for facilitating local application, as well as in indexed international journals. Research infrastructure and research environment, including reward systems Campus-wide effort, with top administrative support, is needed to create infrastructure for successful research including: adequate space for all activities, ICT, library access to literature worldwide, well-equipped laboratories (including skilled maintenance) and field stations. Other elements of ‘enabling research environments’ include institution-wide encouragement and reward for setting strategic priorities, teamwork, taking initiative, productivity, opening opportunities for women, linking research and teaching, transparency in appointment and promotion criteria, efficient procurement, and accountability. As CO looked back in October 2013 from his seat in top management, he reflected that perhaps the single most important characteristic of the research group was the openness and rigour of the two senior members. It is how he saw them working within the University and with collaborators from Sweden and worldwide that may have led to setting new expectations within this group — and soon, perhaps, will be doing so much more broadly. Modern technology is supporting a move towards more open sharing of information — for research and for administrative accountability. As its basis for future support Sida plans to make a new call for proposals, offering funding for specific themes based on Universities’ own future directions for research capacity — in Uganda and elsewhere. The prospects for in-country support for strengthening research capacity at multiple Ugandan universities are growing; however, domestic funds still contribute only a tiny proportion of the resources devoted to research and research training.

Institutions comprise formal and informal rules of conducting business. With increased complexity of global business, the enforcement of formal rules has become exigent. Firms, both domestic and foreign, adhere to these rules to avoid penalty for noncompliance. Since institutions are nation specific, a foreign firm engaging in business in another country will have to be cognizant of the institutions in the host country. The distance between the institutions of the home and host country will therefore influence the decision to do business in the host country and the outcome of that decision. Institutional distance is defined as cross-country differences between multinational corporations (MNCs) home and host countries with regard to their regulatory, normative, and cognitive aspects (Konara & Shirodkar, 2018; Scott, 1995). Extant literature has deemed institutional distance as an important determinant of MNCs' various decisions such as entry mode choice, ownership, staffing, and so on in host countries which in turn influences the performance of foreign subsidiaries (Ambos & Håkanson, 2014; Brouthers, 2002; Chao & Kumar, 2010; Dikova, 2009; Gaur, Delios, & Singh, 2007; Gaur & Lu, 2007; Zaheer & Mosakowski, 1997). Note that though institutions refer to cognitive, normative, and regulatory institutions (Scott, 1995), we confine our study to only the regulatory institutional distance. One reason for that is that regulatory institutions are mostly mandatory and therefore MNCs cannot take them for granted. In other words, they will need to comply with them. Furthermore, regulative institutions change more frequently and rapidly as compared to normative and cognitive institution changes (e.g., cultural changes) (Estrin, Baghdasaryan, & Meyer, 2009). Scholars have also argued that investigating all three institutional distances in a study provides an oversimplified analysis (Zaheer, Schomaker, & Nachum, 2012).

There are different perspectives with regard to how institutional distance has an effect on MNC foreign subsidiaries' performance. On the one hand, the greater level of “liabilities of foreignness” or uncertainty brought by the larger institutional distance leads to greater costs of learning and adapting to the host-country's institutional environment. Thus, foreign subsidiaries would suffer from the declining competitive advantages in the host environment (Shirodkar & Konara, 2017). However, some scholars argue that the greater opportunities for arbitrage such as first mover, and research and development (R&D) advantages due to increasing institutional distance would improve foreign subsidiaries' competitive advantages, resulting in an obvious paradox. An inverted-U shaped relationship between institutional distance and foreign subsidiary performance (e.g., Gaur & Lu, 2007) might be able to explain and resolve this paradox. In addition to the magnitude of institutional distance, the issue of “direction” of institutional distance (i.e., asymmetry) should not be ignored (Contractor, Yang, & Gaur, 2016). MNCs might climb up (when MNCs establish subsidiaries in host countries that are institutionally stronger than their home country) or down (when MNCs establish subsidiaries in host countries that are institutionally weaker than their home country) the institutional ladder (Konara & Shirodkar, 2018). Konara and Shirodkar (2018) argue that climbing down (vs. up) the institutional ladder has a more positive or less negative performance implications of regulatory institutional distance on subsidiary performance and full (vs. partial) subsidiary-ownership is likely to be more beneficial for subsidiary performance in this situation. The current study utilizes institutional theory to investigate one of the MNCs' foreign direct investment (FDI) decisions: R&D, thus adding to the sparse literature in this area. Castellani, Jimenez, and Zanfei (2013) used data on R&D and manufacturing investments of 6,320 companies in 59 countries and found geographic distance has a lower negative impact on the probability of setting up R&D labs than manufacturing plants. They claim that gaining access to the valuable knowledge that is highly concentrated in clusters is the key driver of choosing remote locations for R&D lab. However, institutional proximity (such as the commonality of language, belonging to the same trade area, and sharing similar religious attitudes) plays an important role in this decision process since it influences firms' ability to absorb and transfer knowledge. Higón and Antolín (2012) studied the relationship between R&D and productivity and found that institutional distance between the home and host countries moderates the relationship. In their study, institutional distance is not a determinant of R&D location but a measure of country differences that measures the levels of liability of foreignness local and MNC subsidiaries face respectively. Choi and Contractor (2016) identified a variety of R&D alliance modes in the pharmaceutical industry and found that the likelihood of using a more integrated alliance governance mode decreases as the human capital and cultural distance between nations of the partner firms increases. However, the likelihood increases when the geographic and institutional distances are larger. Obviously, there is little agreement with regard to the determinants of MNCs' R&D investment decisions from the perspective of institutional distance. This study aims to fill this research gap by analyzing the impact of regulatory institutional distance on MNCs' R&D investment decisions. In addition to institutional distance, there are other factors that affect the R&D investment decision. Empirical findings for United States and other MNCs suggest that MNCs prefer to locate their R&D activities in countries that are able to offer them, among other things, larger markets, technological resources, and infrastructure (Kuemmerle, 1999; Shimizutani & Todo, 2008). Host market-oriented affiliates are more likely to have R&D units than the export-oriented ones, especially in developing countries. The relative strength of the patent regime appears to affect the direction rather than the magnitude of R&D investments (Kumar, 1996). These are included as controls in this study. We study the expenditure on R&D by U.S. firms in their affiliates across various countries over a period of 8 years. We find that the regulatory institutional distance affects the probability of R&D investment more significantly than the amount of R&D. Using panel data probit and tobit estimation, we are able to show that regulatory institutional distance affects R&D investment decision negatively and significantly, however it does not have an impact on the amount of R&D. In the next section, we provide the theoretical background and develop the hypotheses. Following that, we describe the data, empirical estimation, and the results. The last section concludes.

2 | THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT 2.1 | Regulatory Institutional Distance Institutions are nation-specific (Kostova, 1997). Institutional distance, which refers to the extent of similarity or dissimilarity between home and host countries' institutions (Kostova, 1997), presents difficulty for a MNC to reap the benefit of internationalization (Dikova, Sahib, & Van Witteloostuijn, 2010). The regulatory environment comprising elements such as constitutions, laws, and property rights, vary in different countries that lead to “regulatory institutional distance” between home and host countries. For instance, formal regulations and rules governing protection of intellectual property rights are more exhaustive, clearly laid out and better enforced in the United States than in China. This creates a regulatory institutional distance between United States and China. Recently, the concept of institutional profile was highlighted to draw researchers' attention to the theoretical and empirical issues present in the extant institutional distance literature (van Hoorn & Maseland, 2016). Institutional profile resonating with the institutional literature at large is the institutional environment of a MNC's home or host country. “Operating in a certain country implies that firms are embedded in and face distinct challenges and opportunities that derive from this country's institutional environment” (van Hoorn & Maseland, 2016, p. 374). Theoretically, institutional profile, different from distance, is not about the dissimilarity between the home and host institutional environments, which creates misunderstanding, legitimacy problems, challenges for knowledge transfer, and organizational routines (Zaheer, 1995). Methodologically, the use of a single country as the reference point when calculating institutional distance between MNCs' home countries and their host countries results in the conflating of institutional distance and profile effects, which should be prevented (van Hoorn & Maseland, 2016).

2.2 | Regulatory Institutional Distance and MNCs' R&D Investment Decisions The environmental complexity MNCs face in foreign countries leads many scholars to conclude that institutional distances have a negative impact on MNCs' internationalization. For examples, Dikova et al. (2010) found the larger the formal and informal institutional distances, the less likely that the cross-border acquisition deal will be completed; Kostova and Zaheer (1999) proposed that the larger the institutional distance, the more difficult it is for MNCs to establish legitimacy in the host country. Furthermore, a recent review suggests that most entry-mode researchers suggested that environmental complexity leads firms to consider partial ownership to maintain flexibility while weathering investment risks (Brouthers & Hennart, 2007). In the same vein, in this study we argue that regulatory institutional distance has a negative impact on MNCs' R&D investment decisions and the amount of the investment, if any. There are two major motivations of MNCs to internationalize R&D (Cantwell, 1995; Kuemmerle, 1997). Traditionally, MNCs conducted R&D activities outside their home countries to support the manufacturing activities of local subsidiaries or to adapt products and technologies developed in their home countries to local market conditions (Doh, 2005). This strategy has been labeled asset exploiting, home base exploiting, or competence exploiting. Asset-exploiting strategies are associated with a view of MNCs as a means to exploit firm-specific advantages in foreign markets (Markusen, 1995; Navaretti & Venables, 2013). R&D by the subsidiaries supports the exploitation by adapting technologies, products and processes to local needs, consumer tastes and regulations, among other characteristics (Dachs & Ebersberger, 2009). However, the home base-exploiting perspective was challenged in more recent years by the observation that MNCs increasingly generate new R&D outside of their home countries. Such a strategy has been described as asset seeking, competence creating, or technology exploring. This motivation for international R&D is to develop new technologies overseas by accessing foreign R&D resources and local technological and scientific strengths with the aim of improving the technological and innovative capacities of the investing firm (Feinberg & Gupta, 2004). The foreign R&D resources can therefore play one or both of two roles: facilitate local adaptation of the MNC's products and services and/or enable the creation and acquisition of globally relevant technology for the entire corporation (De Meyer & Mizushima, 1989). Either motivation requires the compliance responses of MNCs' foreign R&D units to the local institutional requirements, which are critical for them to gain legitimacy in a host market (Raaijmakers, Vermeulen, Meeus, & Zietsma, 2015; Suchman, 1995). For example, in terms of regulatory institutional pressures, the legitimacy requirements are explicitly codified, and usually enforced by a government agency (Scott, 1995). A foreign R&D unit needs to change its practices to comply with the institutional rules to be able to operate legally in the host market (e.g., Chinese firms' compliance with the product safety regulations in the United States). In other words, a large regulatory institutional distance increases the liability of foreignness (Baik, Kang, Kim, & Lee, 2013; Bell, Filatotchev, & Rasheed, 2012). The growing differences between the home and host countries in terms of legal system, the rule of law, and important bureaucratic principles and relationships require MNCs and their foreign R&D units more process adaptation and accompanying documentation, raising the additional cost of doing business in the host market (Valentino, Schmitt, Koch, & Nell, 2019). Quick and safe knowledge transfers are critical for the success of both exploitation and exploration and thus may influence MNCs' R&D investment decisions. According to Valentino et al. (2019), one of the major challenges for MNCs' attempts to transfer knowledge among foreign R&D units is institutional distance. Kostova et al. (2016, p. 14) argued that subsidiaries “from institutionally distant countries are likely to possess and use different information, operate out of different cognitive frames and heuristics, and, as a result, have difficulties understanding and interpreting HQ's (i.e., head quarter's) priorities and requests. This challenges the subsidiary's correct understanding and interpretation of HQ's objectives and requests.” This argument can be applied to the relationship between MNCs and their foreign R&D units, too. Thus, miscommunication and mutual misinterpretation are highly likely to reduce the speed of knowledge transfer from MNCs' R&D investments in the host country to the MNCs or other foreign R&D units. In addition, the different levels of quality with regard to the intellectual property protection between home and host countries will decrease MNCs' confidence on safely transferring knowledge between the MNCs and their foreign R&D units (Contractor et al., 2016). However, only recently did some researchers suggest that deviating from the institutional pressures in a host country may not necessarily put a firm at a disadvantage and, rather, can be an advantage (Shi & Hoskisson, 2012). Deviating from the institutional requirements costs a firm a certain degree of legitimacy, but affords the chance for innovation and exponential growth. For example, Krammer (2018) found MNCs in the global tire industry prefer partners from similar regulatory environments when forming a technological alliance for exploitation and dissimilar regulative partners in exploration alliances. We recognize the potential benefits of regulatory institutional distance on innovation, but we believe those potential benefits will be eventually outplayed by its side effects: miscommunication and mutual misinterpretation. Especially, innovation is usually revealed in the form of tacit knowledge, which is unique to the organization, valuable, usually rare, and thus difficult to be transferred (Rickley & Karim, 2018). The level of difficulty will only be increased when regulatory institutional distance enlarges. Thus, we posit that MNCs are less likely to choose a regulatory institutionally distant country as their R&D location. Even if they do, the amount of R&D investment will be lower (vs. higher) in regulatory institutionally distant (vs. close) country. We formally propose a set of hypotheses as below. Hypothesis 1 Regulatory institutional distance is negatively related to the probability of MNCs' research and development investment in the host country. Hypothesis 2 Regulatory institutional distance is negatively related to the amount of MNCs' research and development investment in the host country.

3 | DATA AND EMPIRICAL ESTIMATION The relationship between R&D investment in an economy and the regulatory distance between the host and home economy can be broken down into the propensity to invest in R&D and the intensity of R&D investment (also known as the extensive and intensive margins respectively). The propensity to invest in R&D in the host country is a binary decision where the dependent variable takes unit value when there is R&D investment in an economy and zero otherwise. The intensity of R&D investment in an economy measures how the amount of R&D investment in the economy depends upon the regulatory quality. The dependent variable in this case is a continuous variable—research expenditure normalized by total sales. The data on multinational firms' research activities come from the U.S. Department of Commerce's Bureau of Economic Analysis (BEA). The firms considered here are majority owned foreign affiliates of U.S. multinationals spread across 164 countries. The data are annual aggregated data on financial structure and operations of the affiliates, and include total sales, total assets, net income value added, R&D expenditure, and employee compensation. For our estimation, we use only the information on R&D expenditure and total sales. Since our data are available over a number of years—2004–2012, we use panel data estimations to investigate this relationship. We use three main independent variables to measure the regulatory distance of the destination markets. The first measure is the World Bank's Doing Business (DB) Report's overall country ranking on regulations, called overall distance to frontier (Overall DTF). The overall ranking is a combination of 11 indicators that evaluate a range of regulations that impact the ease with which business can be conducted in the said country. The index records all official procedures required, and the time and cost to complete these procedures. These indicators are selected by the World Bank based on economic research and entrepreneur surveys from more than 135 economies, which provided data highlighting the main obstacles to business activity. The regulatory distance is measured by the difference between the indices of each of these indicators. The indices range between 0 and 100, with 100 representing the frontier or the highest possible score. These scores provided on an annual basis help assess the absolute level of regulatory performance, the gap between a particular economy's performance and the best performance at any point in time, and how the ranking changes over time. For example, a score of 75 in DB 2010 means an economy was 25 percentage points away from the frontier constructed from the best performances across all economies. A score of 80 in DB 2011 would indicate that the economy improved in that year. Simply put, higher the DTF number better the economy is in that regulatory area.1 The list of the 11 regulatory areas with the countries that ranked the highest or lowest is provided in the Appendix (Table A1). The DB Report has been widely used in various studies (Djankov, Ganser, McLiesh, Ramalho, & Shleifer, 2010; Geginat & Ramalho, 2015). A second indicator of regulatory distance is the measure of Regulatory Quality in the World Governance Index (WGI). This index was developed by Kaufmann, Kraay, and Mastruzzi (2010) and includes a number of regulatory measures, for instance price controls, and burdens imposed by regulation in foreign trade and business development. The index is constructed by isolating the unobserved governance component from each individual data source and optimally combining the data sources to get the best possible signal of regulatory quality. We use the percentile rank of the index that ranges from 100 to 0, with 0 as the lowest rank and 100 as the highest. An important regulation that affects R&D investment is the patent protection regime. As the third indicator of regulatory distance, we use the measure of the strength of the patent protection regime— the Ginarte–Park (GP) index (Ginarte & Park, 1997; Park, 2008). The GP index is the sum of five scores—coverage (industries where inventions are patentable); membership in various international treaties like TRIPS and the Paris Convention; duration of protection; enforcement mechanisms; and provisions for loss of protection (e.g., compulsory licensing and revocation of patents). The GP index has been used extensively in the literature (Claessens & Laeven, 2003; Gooris & Peeters, 2016). We obtained the GP index from Walter Park's website. The index covers 122 countries for the years 1960–2010 and is available at 5-year intervals. It is an unweighted sum of five scores, each of which takes a value between zero and one. Therefore, the GP index varies between zero and five. A higher score represents a stronger patent protection regime. We use the data for the years 2006–2012. Therefore, we have the index available for all countries at two time intervals—2005 and 2010. We interpolate the index for the rest of the years. For our analysis, we consider the years 2004–2012. The time frame is limited by the availability of data. BEA provides data series that is comparable from the year 2004. Though DB Report also provides data from 2004, the combined score, Overall DTF is available for only 5 years, that is, 2010–2014. We have therefore, estimated that specification for the years 2010–2014. WGI's regulatory quality and the Ginarte–Park index are available for the years considered— 2004–2012. The data have not undergone any other transformation or selection. We construct the regulatory distance variables from the above three indices by taking the absolute distance between the United States and the host country. A negative relation between any of these regulatory distance variables and the dependent variable would indicate the importance of having similar regulatory institutions for R&D investment. Table 1 provides the summary statistics of the dependent variable, the regulatory quality indices, and the controls. Table 2 lists the countries that have had the maximum and minimum values for the regulatory quality indicators. Since all three indicators of regulatory quality are correlated with each other (Table 3), we do not combine them in one estimation. We estimate the relationship between R&D and each of these regulatory distances in three separate estimations. The economies are widely dispersed across the above-mentioned indices. The rankings of the economies on each of these indicators also change across time. Some countries like the United States maintain a high score throughout the period of study, whereas others have shown either improvement or deterioration in these regulatory processes. For instance, Hong Kong ranked highest in WGI's regulatory quality index in the years 2004 and 2005. However, in 2011 it did not even make it to the top five ranks. Similarly, Chile was not even in the top five countries in the Ginarte–Park index from 2004 to 2009. However, it ranked second only to United States since 2010. Predictably, some countries like the United States and Canada showed only some change in the regulatory distance. For most other countries, there has been a change in the rankings over a period of time.2 The BEA data did not report the research expenditure of MNCs if it is below USD 500,000 (or 0.5 millions). The BEA also does not report affiliates that are below the exemption level for reporting. The distribution of our dependent variable—the R&D variable—is therefore, censored at the left margin. The true distribution of the R&D variable is the distribution of the latent variable y i that cannot be observed. What is observed in the data is yi which takes a value of zero or more. This can be mathematically expressed as given below.

Since a least square estimation cannot be used to measure this censured distribution, a censored regression model—tobit, with left censoring at zero has been used. The two estimations can then be expressed as given below. The first equation is the reduced form equation for measuring the probability of investing in R&D with the probit model. E½ri,t = 1 > 0 = β0 + β1Xi,t + δt + εi,t: Here, ri,t indicates if the firms invested in R&D in the host country i, in the year t. Since it is a panel probit estimation with a large N and short time period T, the error term εi,t comprises both the individual country effect and the idiosyncratic error. In the above equation, Xi,t measures the regulatory distance for year t. The second equation given below measures the intensity of R&D expenditure (amount of R&D expenditure normalized by amount of sales) with the tobit model. yi,t = α0 + α1Zi,t + μt + ϑi,t: The above equation is observed only if ri,t = 1. Again, the estimation being a panel estimation, the error term ϑi,t comprises the individual country effect and the idiosyncratic error. In the above equation, Zi,t measures the regulatory distance for year t. The amount of R&D investment in the destination country can be driven by the resources available to support R&D. A country that has a large pool of researchers will attract a higher amount of R&D. Following Nielsen and Raswant (2018), we include a number of ontrols—the number of scientific and technical journal articles published by the host country,3 internet users (per hundred), FDI (net inflow) and population.4 The first two are controls for the R&D environment and research resources in the country. The last two control for the size of the economy and the openness of the economy. The data for these variables have been taken from World Development Indicators published by the World Bank. One contention with our estimation strategy could be that R&D investment decision is not independent of the decision to invest in a particular country and our estimation might be picking the effects of FDI location choice. Though we do not claim to establish causality in our model, we do not expect FDI location choice variables to confound our results, as our estimation population comprises of firms that have already made the location choice. Only about 40% of the countries that receive FDI from the United States actually also receive investment in R&D. That means that factors other than those that influence the decision to invest in a country affect the decision to invest in R&D. We use this country level variation in R&D investment to investigate the effect of regulatory distance. To address this further, we introduce a number of controls for FDI choice in our estimation, for instance, net FDI into the destination country, degree of openness, common language, and GDP growth. This estimation strategy might not establish a causal relationship but it does inform the reader about the factors that affect R&D investment in some countries over others. We investigate the relationship at the country level and not at the firm level. One reason for that is that firm level data on country level R&D investment are not publically available. More importantly, there is merit in investigating this relationship at the country level since host country regulatory quality is uniform for all U.S. firms. 4 | RESULTS The estimation of the probability to invest in R&D in a particular country is inversely related to the regulatory distance of the destination country. Columns 1, 3, and 5 of Table 4 shows the probit estimation of this relationship for the three measures of regulatory institutional distance—World Bank's Overall DTF, WGI's regulatory quality and Ginarte–Park's index. All three regulatory institutional distance measures show negative and highly significant coefficients. Columns 2, 4, and 6 of Table 4 show that this relationship remains negative and significant with additional controls. Greater the regulatory institutional distance between the host and home countries, the lower the probability that it would receive R&D investment. Both the number of internet users in the economy and the number of scientific articles published are positive and significant. A country with better research infrastructure would attract more R&D investment. The marginal effects for the three variables are given in Table 5. The margins are calculated at the means as dy/dx and not as elasticities. It can be interpreted as the probability of R&D decreasing with the indicator at a rate such that, if the rate were constant, the probability would decrease by the marginal effects, if the distance increased by one. If the regulatory institutional distance in terms of World Bank's DTF, reduced by one unit, the probability of R&D investment in that country would increase by 1%. The marginal effects are strongest for the patent index. A reduction in the regulatory distance in terms of patent protection would increase the probability of R&D investment by about 18%. Since the regulatory institutional distance is important for the probability of U.S. firms investing in R&D in the host country, we accept Hypothesis 1. The estimation of the amount of R&D investment dependent on regulatory institutional distance gives weaker results. The results of this estimation of the intensity of research expenditure are shown in Table 6. Though all three measures of regulatory quality are still negative and significant, the coefficient is zero in the specifications with overall DTF and WGI's regulatory quality index. The patent regime measure Ginarte–Park index is the only distance that has some influence on the amount of R&D expenditure, albeit a very small effect. Regulatory distance, which influences the probability of getting R&D investment, does not affect the amount of R&D received. Controls for research resources—number of internet users, and number of published scientific articles also have no effect on the amount of R&D investment in the host country.

Based on the above results, we accept the second hypothesis conditionally that the regulatory institutional distance affects the amount of R&D expenditure in the host country only in the case of patent regime, that too to a small magnitude. The other regulatory institutional distances have no effect. 5 | DISCUSSION AND CONCLUSION In this study, we analyze how regulatory distance between countries affects the R&D investment. We believe just like FDI, the institutional distance between host and home countries will have an effect on R&D investment by U.S. MNCs. According to the National Science Foundation's science and engineering indicators,5 Asian countries, most notably China, have heavily contributed to the overall increase in worldwide R&D expenditures. Global R&D shares declined for the United States (37–25%) and for the EU (25–20%) while the economies of East-Southeast and South Asia—including China, Japan, Malaysia, Singapore, South Korea, Taiwan, and India—increased their combined global share from 25 to 42%. This shift in focus on R&D in Asian countries not only increases their share of production and exports of technology intensive industries but also attracts foreign investment in the R&D sector. We investigate if regulatory distance plays a role in the decision to invest in R&D. The data we use for the analysis are the R&D investments of U.S. MNCs, therefore, we measure the regulatory distance of destination countries from the home country, United States. We find support for this theory in our data analysis. Regulations such as rules governing protection of intellectual property rights, competition policy, and market openness policies are exhaustive, clearly documented and better enforced in countries like the United States than in other countries like China. This creates a regulatory institutional distance between United States and the destination countries. Similarly, Bangladesh ranks very low in WGI's regulatory quality index while Indonesia ranks higher. Though both countries receive FDI from U.S. MNCs, Indonesia receives R&D investments for the time period we study and Bangladesh does not. A whole gamut of regulations influences firm activities. Regulations could be in terms of taxation, health, labor and environment standard compliance, getting approvals and licenses and many others. As a result, there is no single measure of regulatory distance. We use three different measures of regulatory distance, one of which—the Ginarte–Park patent index—is more specific to research activities. Countries with explicit regulatory policies have government commitment and the ethos of establishing, implementing, and reforming regulations. Therefore, countries are ranked similarly in all three measures, with slight variation. To address this correlation between the three regulatory distance measures, we estimate the relationship separately. Our results also extend the literature of institutional, specifically, regulatory distance between home and host market. Nielsen, Asmussen, and Weatherall (2017) reviewed and evaluated 153 quantitative studies on FDI location choice from 1976 to 2015 and found factors such as demand, physical infrastructure, human capital, governance/ institutions, special economic zones, industrial cluster, global city, and location experience were the drivers of FDI location choice. With regard to the institutions, 75% (43 studies) find support for the hypothesis that more developed formal institutions attract more FDI (Nielsen et al., 2017). Despite the large attention to “pure economic factors” in these empirical studies, little effort was put into studying their interaction or understanding forces that drive these factors in a particular market. Nielsen et al. (2017) thus called for future research to interact pure economic factors with each other and with local governance institutions. They also found that the application of institutional theory in these studies remains relatively small. This study partially answers the call by Nielsen et al. (2017) for more application of institutional theory in the FDI location studies. Businesses are conditioned to work in the home countries' legal environment. United States has consistently ranked high in terms of regulations such as enforcing contracts. U.S. firms therefore find it easier to work in another country that provides them with similar standards. In the presence of huge costs or procedural delays in various business processes, U.S. firms might not invest in R&D in host economies. For instance, weak enforcement of contracts might increase the appropriability risk of innovations (Zhang, Li, Hitt, & Cui, 2007). Similarly, lower the patent protection in the host country the less likely it is to receive any R&D investment. Our analysis supports our proposition that regulatory institutional distance does negatively affect a foreign firm's probability to invest in R&D in the host country. Though all three measures of regulatory distance significantly affect the probability of R&D investment, they do not have the same effect on the amount of R&D investment. Our article adds to the literature of regulations and R&D activity by providing empirical evidence for majority owned U.S. firms with FDI in 164 countries.

On January 19, 2017, the Obama Administration announced much-anticipated revisions to the Common Rule, a regulation for research involving human subjects adopted by 16 federal agencies (Department of Homeland Security et al. 2017). The Trump Administration has delayed implementation of these revisions until January 21, 2019 (Department of Homeland Security et al. 2018) One of the most significant changes to the Common Rule is the requirement that institutions use a single institutional review board (IRB) for cooperative research in the U.S., unless more than one IRB is required by state, local or tribal law, or a signatory federal agency decides an exception is warranted. The single IRB mandate provision does not become effective until January 20, 2020 (Department of Homeland Security et al 2017) The National Institutes of Health (NIH) has adopted a similar policy for research funded by the agency, which became effective on January 25, 2018 (National Institutes of Health 2016) The primary rationale behind the single IRB mandate is to eliminate redundant review, maintain consistency across different research sites, reduce administrative burdens on investigators and IRBs, and save time and money (Wolinetz and Collins 2017, Silberman and Kahn 2011). Some have also suggested that relying on a single IRB in cooperative research may improve the quality and consistency of review (Wolinetz 2017). Although many investigators, IRBs, and government officials have showed strong support for the single IRB mandate, some have expressed concerns about this new requirement and its implementation (Klitzman et al. 2017). Chief among these are potential difficulties with negotiating contracts (known as reliance agreements or authorizations) to allow one institution to rely on another for IRB review; taking local context issues (such as culture, language, or community values) into account during review; adequate guidance from federal agencies concerning compliance with the mandate; and financial support for implementing the mandate (Wisconsin Consortium 2011, Resnik 2012, Association for American Medical Colleges 2017, Klitzman et al. 2017) The single IRB mandate is potentially the most disruptive change to the Common Rule because it requires institutions to develop policies and procedures for negotiating and implementing reliance agreements, delineating institutional responsibilities, and ensuring that review of cooperative research addresses local context issues. Institutions may need to devote staff time and other resources to deal with reliance agreements and ensure compliance with the regulations. To facilitate reliance agreements, many institutions are using common templates, such as a document developed by the Office of Human Research Protections and one used by institutions that belong to the SMART IRB (Streamlined, Multisite, Accelerated Resources for Trials IRB) consortium (Resnik et al. 2018, Office of Human Research Protections 2011, SMART IRB 2018). The purpose of this study was to survey human research protection program (HRPP) officials at the top U.S. research institutions to understand their knowledge and opinion of the mandate, what steps their institutions are taking, and difficulties their institutions are facing. A secondary goal was to determine whether institutional characteristics, such as the number of human studies reviewed per year, are associated with types of survey responses. Information obtained from this survey may prove useful to investigators, IRBs, HRPP officials, and federal agencies when making decisions concerning compliance with the mandate and best practices for single-IRB review.

Materials and Methods

We obtained a list of the top 200 U.S. research institutions ranked by total research funding from the Center for Measuring University Performance (2014) for the most recent available year, 2014. We eliminated 4 institutions from the list because they were duplicates; 4 because their human research oversight was covered by another (parent) institution; and 4 because they do not conduct research with human subjects. This left us with 188 institutions to contact. We sent an email to HRPP officials at these institutions (e.g. IRB managers, IRB chairs, or IRB office staff charged with overseeing reliance agreements) inviting them to participate in an anonymous online survey. The email included text needed to obtain informed consent (e.g. description of the study, benefits, risks, voluntary nature of participation, etc.) and included a link to the survey. The survey consisted of 13 closed-ended and open-ended questions (see Supplementary Material). We sent two reminder emails. For data analysis, we presented descriptive statistics by count and percentage. We used Spearman’s correlation test and Fisher’s exact test to assess the association between responses to survey questions and institutional characteristics. The NIH Office for Human Subjects Research Protections determined that our study qualified as exempt research. 107 institutions (56.9%) responded to the survey, which is much higher than the typical response rate of 33% for online surveys (Nulty 2008). An overwhelming majority (96.3%) of the respondents were universities, medical schools or academic medical centers. 73.8% were public, and 67% review 200 or more human research protocols each year. See Table 1. The majority of respondents (67.1%) rated their knowledge of the single IRB mandate as excellent or very good. Approval of the mandate was mixed, however. Slightly more respondents (38.6%) strongly approve or approve of the mandate than those who strongly disapprove or disapprove of it (34%), and 27.4% neither approve nor disapprove of it. Most respondents agreed that the mandate is likely to have some benefits, including reducing redundant IRB review (80.1%), encouraging cooperative research (46.2%), saving money or resources (27.9%), and enhancing the quality or consistency of IRB review (22.1%). Respondents also recognized, however, that the mandate could create some difficulties, including problems with negotiating reliance agreements (71.2%); reviewing the local research context (68.2%), legal liability (68.2%); securing adequate resources to comply with the mandate (65.4%); educating or training investigators, IRB members, or IRB support staff 69 (64.5%); implementing reliance agreements (63.6%); and developing policies or standard operating procedures (46.7%). See Table 2. Although the quantitative data indicate that most respondents approve of the mandate, some of the comments we received expressed strong disapproval for it or stated that their institution is likely to have serious problems complying with it. In the responses to the open-ended questions 25 addressed concerns related to addressing the local context when a single IRB reviews research, 14 mentioned problems with securing adequate resources (such as HRPP staff) needed to negotiate and implement reliance agreements, and 9 suggested that institutions may outsource human subjects review to commercial IRBs in some cases to comply with the mandate. (Text of comments available upon request.)

Respondents said their institutions were taking various steps to comply with the mandate, including developing policies or standard operating procedures (81.2%); educating or training investigators, IRB members, or staff (77.1%); negotiating reliance agreements with other institutions (60.0%); developing or acquiring reliance agreement templates (58.1%); and allocating resources to comply with mandate (43.8%). 11 respondents who responded “other” to the question related to complying with the mandate indicated that their institution was joining the SMART IRB consortium. 35.5% of respondents said they expected their institution to have no difficulty or minimal difficulty complying with the mandate; 29.9% said they expected their institution to have more than minimal but not significant difficulty complying; and 34.6% said they expected their institution to have significant or serious difficulty complying. 46.7% of respondents said that their institution has faced difficulties with negotiating or implementing reliance agreements in the past five years, but only 13.3% said they had decided not to engage in cooperative research because of such problems. Respondents from larger institutions tended to have a better knowledge of the mandate (correlation coefficient, r = 0.19; p-value = 0.053) but also a less favorable opinion of the mandate (r = −0.18, p-value = 0.074) than those from smaller ones. Respondents from larger institutions were also more likely to have had difficulties with negotiating reliance agreements (r = 0.21; p-value = 0.035). Respondents from medical schools, academic medical centers and institutions in the “other” category tended to have better knowledge of the single IRB mandate compared to colleges or universities (Fisher’s exact test p=0.016). Respondents from private institutions were more likely to decide not to engage in cooperative research due to problems with reliance agreements (r = −0.21, p=0.034). (See Tables 3–7). Discussion

While support for the single-IRB mandate was positive overall, most respondents acknowledged that their institution is likely to face some difficulties complying with it, such as those related to reviewing the local context, dealing with legal liability issues, securing resources, education/training, and negotiating or implementing reliance agreements. Respondents from larger institutions tended to have a less favorable opinion of the mandate than those from smaller ones, perhaps because they anticipate that their institution will have more difficulty related to negotiating or implementing reliance agreements required by the mandate, securing adequate resources for compliance, or being required to be the institution responsible for the IRB review because of its size. Adopting common reliance agreement templates could help minimize problems with negotiating these agreements but not necessarily those related to implementing them, since implementation requires collaborating institutions to share information pertaining to research review and oversight and to meet their respective responsibilities. It remains to be seen whether the single-IRB mandate will have its intended effect of reducing administrative burdens. Drawing on the quantitative data from this survey as well as the open-ended comments made by participants, it is reasonable to infer that the mandate may simply shift administrative burdens from the IRB to the HRPP and research staff as they deal with complexities related to reliance agreements, local context, and legal liability. Comments made by some of the survey respondents express these concerns: Although only one IRB may review, the workload on IRB Administrative staff increases because of the “mandate”. Local administrative view and work to coordinate and negotiate reliance agreement remains, and now can no longer handle each in ways that are most efficient given the context of the project (Respondent 1). On the few done so far, it is a tremendous amount of work for the IRB office and has been very, very confusing to determine what is covered by whom for the first large cooperative/collaborative research venture which involved databank samples. It was much simpler just to do a one-on-one IIA than undertake this sIRB model (Respondent 2). I think it is important for institutions to realize that while the IRB Committee workload may decrease when relying on another IRB, the institutional workload most likely will increase. When serving as the Reviewing IRB, the IRB Staff and Expedited Reviewer workloads have the potential to increase significantly (Respondent 3).Others were very concerned about securing funding for increases in HRPP staff needed to comply with the mandate: NIH required this without providing the funding to implement it. We need at least 2 staff members to do all this additional work related to reliance agreements, legal reviews, and requests to rely, yet no funding is available. People expect a quick turnaround but the administrative reviews for other IRBs still take time. And the ability to charge for the reviews is quite complex (Respondent 4). [We are] creating new budget models to assist PIs with ascertaining costs associated with the use of external (commercial) IRBs in the event that our institution does not choose to be IRB of Record. We are not only “allocating resources,” we needed to seek new funding to cover the new work within the IRB office (Respondent 5). Reliance agreements can take a significant amount of time to negotiate (including educating PIs about their responsibilities, etc.) and we don’t have the staff/resources to deal with these agreements on a large scale (Respondent 6).Some respondents speculated that the increased costs associated with mandate will lead institutions to outsource IRB review to private companies: Ethical reviews may be outsourced to commercial IRBs because of fear of liability or lack of institutional resources (Respondent 7). Counter-intuitively, I’ve had investigators design research to NOT be cooperative because they don’t want to deal with sIRB, especially since my IRB is outsourcing the sIRB review to a commercial IRB, so investigators don’t want to have to learn the ins and outs of a whole new IRB (Respondent 8). [We] have spent a lot of time and energy determining how to track sIRB protocols since we are relying on a commercial IRB to conduct all of our sIRB reviews and we don’t have great communication among the IRB, grants and contracts, and investigators (Respondent 9). A potential limitation of our study is that our sample was small (107 responding institutions) and focused on the top 200 U.S. research institutions. Such institutions may have more resources which may influence perceptions regarding the inclusion of single IRBs. However, the strength of this survey lies in the specializations of the individuals that were included (i.e. HRPP officials) who are likely to have excellent or very good knowledge of the mandate and the steps their institutions are taking to comply with it. Another potential limitation is our survey addressed general concerns with the single IRB mandate and did not include detailed questions concerning institutions’ specific experiences with collaborative research involving review by a single IRB. While it would be useful to obtain this information, we chose not to solicit it in order to keep our survey short enough to promote a high response rate. Follow up studies related to specific experiences with single IRB review could help to provide data that would be useful to policymakers. Since the recent modifications to Common Rule implies the broad use of single IRBs in cooperative research, this area of study will become timely and necessary in order to understand benefits and challenges of this new model. In conclusion, our survey indicates that the single-IRB mandate is likely to have both positive and negative impacts on ethical and legal oversight of research with human subjects. Regulatory agencies, such as OHRP and NIH, can help institutions to comply with the mandate by providing guidance concerning such issues as exceptions to the mandate, local context review, oversight and implementation of reliance agreements, and development of policies, procedures, and best practices. It remains to be seen whether the single IRB mandate will reduce administrative burdens without compromising protection of human subjects. Fortunately, the mandate does not become effective until early in 2020. Between now and then there is still much work to do.

Supplementary Material

176×

This paper describes a number of approaches which can be used to analyse the changes taking place within and amongst universities subject to a Performance-Based Research Funding System (PBRFS). Since first implemented in the UK in 1986, there has been a proliferation of these schemes. Hicks (2012, p. 251) observed that the widespread adoption of these systems represents a shift in national research policy since universities are central in national innovation systems. 1 There is nevertheless, geographical variation in the use of these schemes. For example, their use to fund university research has become more widespread in Europe, Oceania and Asian countries such as South Korea and Hong Kong but is hardly used in the USA (see Jonkers & Zacharewicz, 2016). Schemes vary in their details but are all designed to ensure accountability and incentivise performance. Hicks (2012) identified several common characteristics of PBRFSs. They are as follows: Research output must be evaluated; the evaluation must be ex post; government distribution of research funding must depend on the results of the evaluation and the scheme must be national. Variations in schemes can arise from the range of outputs that are used to assess performance and funding. Some schemes, as in the UK, focus only on selected outputs, whilst others, such as the New Zealand and Italian schemes, focus on all research outputs. Schemes may be based on peer review, bibliometric data or a combination of these (Geuna & Martin, 2003; Hicks, 2012; Kolarz, Arnold, Dijkstal, Nielsen, & Farla, 2019 and OECD, 2010). In a wide-ranging review, de Boer et al. (2015, p. 5) suggested that evidence on their effects was limited. More recently, Hicks (2020, p. 352) remarked that “the question of what, if any, are the effects of these policies has only become more urgent.” Nevertheless, there has been increased research momentum evaluating the impact of PBRFSs. The issues investigated have been diverse. There has been research investigating whether they have: resulted in a higher volume of research being achieved at the expense of quality (Butler, 2003; Checchi, Malgarini, & Sarlo, 2019; Hodder & Hodder, 2010; van den Besselaar, Heyman, & Sandstrom, 2017); changed the location of publications resulting in local researchers seeking to increase publications in overseas English-language journals (Mathies, Kivisto, & Birnbaum, 2020); discouraged publication in second-tier journals which serve important training and knowledge-bridging roles and publish topics not well served by mainstream journals (Chavarro, Tang, & Ràfols, 2016); and influenced the extent of self-citation (Abramo, D'Angelo, & Grilli, 2021). There has also been a considerable enquiry into attitudes and management responses within universities. For example, Lewis (2014) assessed the attitudes of academics to these schemes, and whether those attitudes affect research performance. Cattaneo, Meoli, and Signori (2016) have investigated whether the degree of social acceptance of a university has an influence on the response of a university to the introduction of a PBRFS. The effect of PBRFSs on university governance and management practice and the possibility of “gaming” responses has been another area of enquiry; see Woelert (2015, 2021). Abramo, D'Angelo, and Di Costa (2019) enquired whether there has been a tendency for the use of “gift authorship” to enhance the research publication record of academics with previously poor research performances. Cost and benefit analysis of these schemes has been another area of investigation; see Geuna and Piolatto (2016) and Hazledine and Kurniawan (2005). A fundamental issue in evaluating the impact of a PBRFS is the impact these schemes have on the research performance of universities and how universities respond to the incentives created. The aim of this paper is to propose and illustrate the application of a variety of analytical methods to assess the impact of PBRFSs on the research quality of universities. A detailed discussion of what constitutes high-quality research, though this has been much debated, is beyond the scope of the paper, which is to present methods to examine, in a social accounting framework, the transitions and flows that follow from the incentives created by a PBRFS. These methods derive from the fundamental point that, whatever the form taken by a PBRFS, changes in the measured research quality of a unit arise from a limited number of factors. First, there is the turnover of staff, involving the entry and exit of researchers. Second, there are quality changes, or “transformations”, of individuals who remain in the unit. Changes may arise from changes in the composition of multi-disciplinary units, whereby lower performing disciplines are contracted, whilst higher performing disciplines are expanded. Accordingly, these methods can be applied to schemes that assess individual researchers, research groups or a selection of research groups. In view of the central role of these flows, the analytical methods presented here involve the use of a social accounting framework. This enables the components of change—the exits, entrants and quality transformations—to be isolated. The methods can be applied to many different types of PBRFS and are amenable particularly to schemes which maintain records for individual researchers, although in some cases, longitudinal information about more aggregative units of assessment can be used. In the case of the REF in the UK, where individual records are not maintained and where only a proportion of staff are assessed, Stern (2016, p. 20) recommended that all research-active staff should be examined. A problem in attempting to evaluate the effects of PBRFSs is that information about the metrics used is not typically available before the schemes are implemented, and there are no control groups available for comparison. Of course, this problem arises with many new policies where the randomised selection of control groups is not possible, or natural experiments do not exist. In the absence of this kind of information, some commentators have argued that observed changes would have taken place in the absence of a PBRFS, given a normal process of competition amongst universities, and the existence of external factors influencing change: on these difficulties, see Hicks (2017). One approach is to use a bibliometric index of research output as a proxy for research performance and to examine how it compares before and after the introduction of a PBRFS; see Butler (2003), Adams and Gurney (2010), Anderson and Tressler (2014), van den Besselaar et al. (2017) and Abramo and D'Angelo (2022). Wang and Hicks (2013) use time-series techniques to search for a structural break in a bibliometric index for the UK and compare the estimated timing of a break with “narratives of policy change.” Similar time-series methods are used by Tonta (2018) to evaluate the impact of the introduction of a PBRFS on a bibliometric measure of the research performance of universities in Turkey. Checchi et al. (2019) also use narratives to judge the timing of a policy-change dummy variable in a regression analysis. Bibliometric indicators have also been applied using a production function approach; see Moore, Newman, Sloane, and Steely (2002) for the UK and Smart (2009) and Gemmell, Nolan, and Scobie (2017) for New Zealand.2 The question of whether there is a strong concordance between the bibliometric indices and the type of output that PBRFSs measure is addressed by Ancaiani et al. (2015), Abramo et al. (2011) and Baccini and De Nicolao (2016) in the context of the Italian research assessment exercise and by Aksnes and Taxt (2004) for Norway; see also the review by Aksnes et al. (2019). Another approach to evaluation, suggested here, is to consider explicitly the nature of the incentives generated by the PBRFS and their implications for the fundamental flows mentioned above. This provides an understanding of the precise nature of the responses by institutions subject to new incentives. This point is recognised by Woelert and McKenzie (2018) who evaluate how universities in Australia modified their internal performance assessments to align with the financial incentives in the Australian PBRFS. When these incentives are delineated, along with constraints imposed on university managers, it is possible to investigate whether changes are consistent with those incentives. To illustrate how the incentives can be related to the flows taking place within the social accounting framework, it is necessary to understand the details of the metrics used. To provide a context, the methods are illustrated using a data set collected as part of the New Zealand PBRFS. The metrics and associated incentives are explained in Section 2. An understanding of the evaluation method is crucial to appreciate the incentives created. Furthermore, the special nature of a scheme indicates the features of staff turnover to which special attention needs to be given. Whilst any approach cannot provide conclusive proof, the identification of substantive changes that are consistent with the incentives provides prima facie evidence of a policy effect. Section 3 explains how changes in research quality and the entry, exit and quality transformation of researchers can be traced using the social accounting framework. Section 4 applies a decomposition method to identify the separate contribution of these components. Section 5 shows how the contribution of changes in discipline composition can be examined. The question of whether a PBRFS results in different types of convergence is examined in Section 6: this method does not require individual data and can be applied to any PBRFS that publishes a quality measure of units of interest. Section 7 shows how the possible sustainability of changes can be examined by looking at longer run projections. Section 8 concludes. 2 | A PBRFS AND CHANGING INCENTIVES The approach taken in this paper is motivated by the fact that PBRFSs were introduced with the explicit aim of providing incentives to universities. Those incentives are embodied in the precise metrics used and the consequent allocation of public funds. This section shows how the assessment method used in NZ can be linked directly to the flows mentioned above. 2.1 | The NZ PBRFS and measurement of research quality The essential features of the NZ system have remained the same for all assessment rounds, in 2003, 2012 and 2018.3 Three measures are used to allocate funding. The largest component is Quality Evaluation (QE) and represented 60% of total PBRF funding until 2018, when it was reduced to 55%. QE assessment is based on the performance of all eligible researchers over the previous 6 years. The other two components are a subsidy received for each dollar of External Research Income raised (this represented 15% of PBRF funding until 2018, when it was raised to 20%), and a subsidy for Research Degree Completions, which represents 25% of PBRF funding. Funding from these components represents about 10% of the total revenue of New Zealand universities. The present paper concentrates on the largest component, the Quality Assessment.

The Quality Evaluation produces a quantitative performance score, referred to as an Average Quality Score (AQS), for each subject area and university. Each researcher, h, is assigned to a quality category (QC), as determined by a panel of assessors.5 There are four QCs, indicated by A, B, C and R, where the highest category is A, and R indicates an absence of significant research outputs.6 Each individual is given a score, Gh, depending on the letter grade: 10 for A; 6 for B; 2 for C and 0 for R. The AQS is the employment-weighted arithmetic mean score, obtained as follows: If the full-time equivalent (FTE) employment weight of person, h, is eh ≤1, and n denotes the number of employees in the group, the AQS is:

The data used here include the QC assigned to every researcher who participated in any of the three full assessment rounds. Those eligible include all research and teaching staff who are employed on the PBRF census date under an employment agreement with a duration of at least 1 year and are employed throughout the contract on at least a 0.20 FTE basis. Hence, there is little incentive for universities to switch low research-intensive academics to more intensive teaching and administrative roles to avoid inclusion in the PBRF, and although precise numbers are not available, there does not appear to have been a significant tendency to make shifts of this kind. 2.2 | Incentives and responses in the NZ system Universities were obviously incentivised to recruit higher quality researchers and substitute them for lower quality researchers. The new funding system linked a dollar amount of research funding per researcher to each university per “point” in the numerical score, G. Hence, for example, the funds received for an A-rated researcher (for whom Gh = 10) were five times the amount received for a C-rated researcher (Gh = 2). However, the incentives vary amongst universities depending on their initial AQS. For example, the proportional increase in the AQS of a university, from hiring one additional type-A researcher, depends on its initial AQS and size. By contrast, the change in AQS from eliminating type-R people depends only on the size of the university and the reduction in Rs. If a university has an AQS greater than 2, its AQS would fall as a result of hiring one more C-type researcher. It would need to recruit at least at level B or reduce the number of R-type people and substitute them with a new C or higher quality researcher. There are sharply decreasing returns from hiring A and B researchers. These properties suggest that responses to the PBRFS are likely to vary depending on the university's initial situation. Given these features, the expected exits, entrants and quality transformations are as follows: There is an unambiguous incentive to minimise the number of Rs. Exits of Rs do not need to rely on normal retirement, since the nature of contracts means that it is possible (at a cost) to “manage out” those who are not performing at required levels. However, in view of the multiple outputs of a university, involving teaching and external engagement as well as research, an appropriate staffing balance is obviously required.

In 2003, some universities and disciplines had AQSs below 2 (the value of Gh for Cs). Hence, in those cases, there was an incentive to recruit Cs. In 2012, the AQS for most universities and disciplines exceeded 2, so the option of reducing the proportion of C's is consistent with increasing the AQS for all universities. Furthermore, a rise in the entry rate for As and Bs, compared with that for Cs (and Rs), is likely for the 2012–2018 period. By 2012, with a higher proportion of researchers in QC categories with G > AQS, there is a stronger incentive to increase the sum of A and B researchers as a proportion of total researchers by 2018, compared to 2003 and 2012. The time and resources required to convert an R researcher into an A-quality researcher are likely to exceed those of converting a B or C researcher to an A. Similarly, the transformation rate of Bs to As is expected to be higher than for Cs to As. Differences amongst disciplines in the level and rate of improvement are likely to arise from differences in research methods, funding opportunities, alternative labour market opportunities and the opportunity costs of careers in academia and research; see Ehrenberg, Kasper, and Rees (1991), Boyle (2008) and Xu (2008). Similarly, research productivity is influenced by individual and institutional characteristics, research methods and availability of contestable research funding, which vary amongst disciplines; see Wanner, Lewis, and Gregorio (1981), Shin and Cummings (2010), Jung (2012) and Sabharwal (2013). 3 | MAPPING RESEARCHER DYNAMICS For illustrative purposes, Table 1 reports the transition proportions (the flows as a proportion of the initial total number in each category) for all universities combined, for movements from 2003 to 2012 and from 2012 to 2018. The flows are from rows to columns. Consistent with the PBRFS incentives, the largest exit rate during the period from 2003 to 2012, at just over 70%, is of those assessed as R-quality researchers in 2003. The exit rate of Cs from 2012 to 2018, at just over 39%, was higher than for As and Bs, which again is consistent with the incentives. Exits of Cs were just under 48% of total exits. A higher entry rate for As, Bs and Cs from 2003 to 2012 compared to that for Rs can also be regarded as a response to the new incentives. The highest entry rate is for Cs, at nearly 52%. This is consistent with the incentive to substitute new higher quality researchers for the high rate of exit of Rs that occurred from 2003 to 2012. However, there is no tendency for the entry rate of As and Bs to rise from 2012 to 2018 compared to the entry rate for Cs, which remained the dominant source of new entrants from 2012 to 2018. This reflects the relative scarcity of higher quality researchers and contributes to a slower rate of improvement in AQSs during the second period. A large change involves the proportion of Rs over the period, which fell from 34% in 2003 to about 5% in 2012 and less than 3% in 2018, with exit rates over the two periods of 70 and 54%. The proportion of As and Bs rose over both periods, again as expected. The transformation rate of Bs to As was about 18% in both periods; for Cs, it was close to 3% from 2003 to 2012 and 2% from 2012 to 2018. For Rs, the transformation rate was negligible. The same pattern is evident for rates of quality transformation of Cs and Rs to Bs. As expected, the transformation rate of Cs and Rs increases as the target QC is reduced. The statistical significance of the differences in observed transition rates between the two periods can be examined using standard χ2 tests, calculated by comparing transitions observed for 2012–2018 to those expected using observed rates for 2003–2012. This approach is particularly insightful when applied to individual universities and discipline groups.7 In the NZ context, universities and disciplines differed significantly from each other in ways consistent with the incentive effects discussed in Section 3 (see Buckle, Creedy, & Gemmell, 2021). Part of these differences arose from the fact that their initial level of research quality differed. In particular, the initially lower quality universities and disciplines were able to recruit mid-quality researchers and still improve their overall quality of research. Because they initially had a larger proportion of low-quality researchers, substantial overall improvements were possible by encouraging the exits of those researchers. For the initially higher quality units, it was necessary to recruit a higher proportion of higher quality researchers. These differential responses, consistent with the incentive structure, also contribute to the differential growth rates discussed in Section 6. 4 | DECOMPOSING CONTRIBUTIONS OF ENTRY, EXIT AND QUALITY TRANSFORMATION The social accounting framework can also be used to understand the separate contributions of exits, entrants and quality transformation to improvements in AQSs. The characteristics of these components contribute to differential responses consistent with the PBRFS's incentives. The decomposition can be achieved by suitably modifying the relevant transition matrix. Let, the initial and final AQS for the specified group be denoted by Q1 and Q2. An AQS can be calculated for the final period, using an assumption that there are no entrants into any category; this is denoted by Q3. Furthermore, an AQS in the final period can be obtained by setting all exits and entrants to zero; this is denoted Q4. The assumption of no exits is applied by supposing that those recorded as exiting remain in the quality category in which they were placed in the initial period, that is, the diagonals of the flow matrix are augmented by the number of exits. The difference,

Q4-Q1, therefore reflects the effect of quality transformations made by those who remain in the system. The difference, Q3-Q4, reflects the separate effect of exits. Finally, the difference, Q2-Q3, measures the effect of entrants. These are combined to give:

Q2-Q1=(Q2-Q3)+(Q3-Q4)∔(Q4-Q1)

These annual average contributions, for all universities combined, are summarised in Figure 1. The components, for each university (all disciplines), and discipline group (all universities) can be obtained in the same way (see Buckle, Creedy, & Gemmell, 2021). The approach can be used to show that there was greater heterogeneity in the contributions of exits, entrants and quality transformations to changes in AQSs for disciplines within universities. The net impact of exits on AQSs was positive for all universities and discipline groups in both periods; the net impact of entrants is always negative and the net impact of quality transformations is always positive. The positive contribution from exits was much larger from 2003 to 2012 than the subsequent period. The negative contribution from entrants was larger in the second period, and the contribution from quality transformations was larger in the second period. During the first period, the largest proportion of exits was of Rs. During the second period, the largest proportion of exits was of Cs, and there were more B and A exits than from 2003 to 2012. As the quality of incumbent researchers improves, the quality of exiting researchers rises, for a given turnover rate, and again the exit pattern is consistent with the incentives. The negative effect of entrants means that the average research quality of entrants was lower than the end-of-period AQS of those researchers who remained in the system. The AQS of entrants can nevertheless be above the AQS at the beginning of the period. In the second period, the higher initial AQS (in 2012) made it more difficult to recruit entrants above the average quality. Unless entrants have an AQS greater than or equal to the final AQS of incumbents, they cause the AQS to fall. The net impact of transformations is consistently positive, reflecting decisions to attract and retain researchers likely to improve. The same approach can be used to reveal that there are significant differences between universities and disciplines in the range of the components. For universities, the largest range in both periods was in the contribution of exits to improved AQSs. 5 | CONTRIBUTIONS OF RESEARCHER QUALITY AND DISCIPLINE COMPOSITION TO AQS CHANGES One further way for a university to improve its AQS is to exploit differences amongst discipline groups by changing the discipline composition of researchers.8 This section shows how the extent to which AQS changes can be attributed to a changing discipline composition, compared with changing quality, of researchers. The method follows Buckle and Creedy (2020, p. 120), which uses the type of decomposition suggested by Shorrocks (2013) and is initially used to examine the contribution of factors, such as population structure, to changes in income inequality.

The first term reflects the change in AQS attributed to the changing quality of staff, given the discipline composition in period 0. The second term reflects the change in AQS attributed to the changing discipline composition, given the quality of staff in period 1. There is no special presumption in favour of using (4) or (5). Hence, one approach is to take the arithmetic mean of the components. For illustrating purposes, Table 2 reports absolute changes to the AQS attributed to changes in research quality and discipline composition (in FTEs) for each NZ university. From 2003 to 2012, all the increases in the AQS, for all universities combined, came from quality changes. Amongst universities, there are some small contributions from discipline composition change. From 2012 to 2018, the contribution from quality improvement was smaller than in the previous period but remained the dominant source of improvement. In one case, the contribution from composition change is negative. Hence, NZ universities generally did not systematically exploit strategies involving reductions in the size of departments. One exception was the education discipline group, which had the lowest AQS in all rounds. Teaching commitments and other institutional constraints prevent substantial variations in staff/student ratios. For example, in NZ, programmes must be approved by the Vice-Chancellors' Committee on University Academic Programmes, and the TEC determines which programmes receive government funding. 6 | INCENTIVES AND DIFFERENTIAL GROWTH RATES OF AVERAGE RESEARCH QUALITY This section examines how the design of a PBRFS can lead to systematic differences in growth rates of research quality for universities and disciplines. Subsection 6.1 discusses, in general terms, the paths by which differential growth arises. Changes that result in a “catching up” process, whereby low-scoring universities or disciplines on average have higher growth rates than initially higher scoring universities, are called β-convergence. Reductions in the variability in research quality are called σ-convergence. This terminology is taken from the cross-country growth literature; see Barro and Sala-i-Martin (1991) and Quah (1993). Section 6.2 presents an empirical specification that can be used to test for convergence.

6.1 | Growth and incentives An inevitable consequence of a PBRFS is that the profile of assessed research quality of a discipline area or university over time eventually displays decreasing rates of increase. This arises regardless of the method of assessment, the metrics used to measure quality or the allocation of research funding. It arises fundamentally because of the fact that the research quality of a unit can be raised only by quality transformations (of existing researchers who remain in the unit) and by turnover. At low-quality levels, growth may be rapid in view of the greater ability, at lower cost, to encourage the exit of low-quality researchers, and the ability to recruit mediumquality researchers who nevertheless contribute to growth, even though it may be difficult to attract high-quality researchers. If a greater proportion of researchers are of medium or high quality, improvement can arise only from recruiting rarer, more expensive, high-quality research. There is an upper limit, characterised by the case where all researchers in the unit are judged to be of the highest quality, and financial and other constraints impose a limit to the size of the unit. The precise shape of the growth profile depends on the cost and ability to attract new highquality researchers, the flexibility to manage the exits of low-quality researchers and the ability to encourage “stayers” to improve their research quality. There may be constraints on the ability of some units to pass beyond a certain stage of the profile, in view of scarcity and cost considerations. Although there is in principle an upper limit to the quality of a research unit, this does not necessarily imply that a catch-up process of β-convergence operates. The shape of the profile, and ease of upward movement, depends also on the precise nature of the incentives in its choice of metrics and financial rewards. Consider a situation in which the assessment process gives credit only for research judged to be of high quality, above some threshold level. Whether the overall quality is based on some average of individuals or on a unit-level assessment is not relevant. It means that—given the fundamental role of staff turnover—encouraging the exit of lower quality researchers, recruiting medium-quality researchers, or generating quality transformations, may have little effect on the overall evaluation of the unit. In addition, high-quality researchers in a low-quality unit may find that conditions deteriorate and they are encouraged to move. Attempting to recruit high-quality researchers may be difficult, costly if a sufficient threshold for the unit is not achieved. In this situation, low-quality units may face a strong barrier to growth. Even though the units at the high-quality end of the range face a limit to growth, there may not be a significant catch-up process. With increasingly more resources going to high-quality units, and the difficulty facing low-quality units, there may be divergence or polarisation of research units. The precise extent depends on the nature of the academic labour market (including the extent of outside employment opportunities) and the nature of the assessment and resulting allocation of resources. Some funding schemes have led to the channelling of resources to the bestperforming institutions and disciplines, resulting in a degree of specialisation and concentration; see Hare (2003), Barker (2007) and Broadbent (2010). However, β-convergence may arise where the assessment does give credit for lower-tomiddle research quality within units and uses some kind of average performance rather than looking only for evidence of outputs judged to be above a threshold quality level. In such a structure, it is possible for the low-quality units to use turnover to raise their overall assessment. They may, therefore, also be able to retain their existing high-quality researchers. Under these conditions, catch-up can take place. In view of the nature of the New Zealand PBRFS, such convergence seems likely.

6.2 | Testing for convergence This section presents a specification, following Buckle et al. (2020, pp. 3,923–3,924), which can be tested, using the quality scores for universities and discipline groups.9 First, consider discipline areas, and let qi,j denote the AQS for discipline i in university j. Between two rounds, these change by absolute amounts Δqi,j. Assuming that all quality changes are positive, convergence amongst any two disciplines, i and k, within a university, j, exists when, for qi,j <qk,j:

That is, the proportional growth in measured average quality is larger for those starting from a relatively lower initial average quality. For convergence amongst a wide range of disciplines within a university, it is required only that there is a significant tendency for the relatively weaker disciplines to improve proportionately more than the stronger disciplines. Thus, consider the following specification (for variations in i for given j), where an additional subscript relates to the time period:

Here, u is a Normally distributed random error term, and αj is equal to μj,t 1þβj μj,t1, where μj,t is the (unweighted) logarithm of geometric mean quality in university j at time, t. This arises from a process in which proportional changes in quality, relative to the geometric mean, are related to the initial relative quality. A value of βj equal to zero implies that all proportional changes, log qi,j,t log qi,j,t1, are equal to the proportional change in geometric mean quality, except for the stochastic variation. The extent to which βj is less than zero measures the degree of systematic β-convergence.10 With β-divergence, the addition of disequalising changes via the random component, u, ensures that the dispersion of q increases from t-1 to t. However, β-convergence does not itself ensure σ-convergence, which refers to the change in the distribution of quality scores. Denoting the variance of ui,j,t in (7) by σ2 u, the change in the variance of log qi,j,t is given by (dropping i,j subscripts):

The term on the left-hand side captures σ-convergence if it is negative or σ-divergence if it is positive. Clearly, σ-convergence cannot arise with β-divergence (β > 0). However, given β-convergence, σ-convergence requires σ2 u=σ2 q,t1 < βð2þβÞ. When using data for all disciplines and universities, a regression equation of the following form, based on (7), can be used:

Differences across universities and disciplines in the growth and convergence or divergence of their AQS values can be examined using shift dummy variables, Di and Dj, where D = 1 for each university or discipline in question and zero otherwise, and similar slope dummies, Di logAQSit1 ð Þ, and Dj logAQSjt1 . The parameters α and β and their university and discipline-specific equivalents therefore respectively capture autonomous AQS growth and the rate of convergence (β <0) or divergence (β>0) of AQSs for each university and discipline. The parameter, β, may be interpreted as the rate of convergence conditional on autonomous sources of AQS growth, including any university or discipline fixed effects. The dummy variables can be progressively eliminated using a “general-to-specific” or “specific to general” approach; see Castle, Doornik, and Hendry (2011) and Hendry and Doornik (2014). Convergence tests can also be conditioned on a vector of additional variables. Using NZ data, Figure 2 shows logAQSij2012logAQSij2003, converted to annual values by dividing by 9, plotted against logAQSij2003. The figure shows 87 observations, for 8 universities, 9 broad discipline groups and 70 disciplines within universities (one university has only 7 disciplines). There is a close relationship between initial AQS and subsequent AQS growth, reflecting strong β-convergence. Where three separate years' are available, an additional lagged term can be added, giving:

logAQSijt \_logAQSijt\_1 .γtβ1 logAQSijt\_1tβ2logAQSijt\_2tuijt

This specification is also consistent with one having the logarithm of the t-1 AQS, and the growth from t-2 to t-1 (the difference in logarithms of AQSs), on the right-hand side. The full convergence effect, β, is measured by β = (β1 + β2). The results of estimating Equation (10) for NZ are given in Table 3, showing substantial β-convergence, with β1 = 0.1167.11 The only significant dummy variable is the slope dummy relating to the discipline group, “Accounting, Finance and Economics” (AFE). The inclusion of logAQS2003 slightly reduces the longer run annual convergence rate; with β2 = 0.022 and hence β = 0.0947. The positive sign on logAQS2003 reflects the fact that the degree of convergence in the first period is negatively associated with convergence in the second period, thereby modifying the full-period degree of convergence. This slightly slower rate of convergence when estimated over the whole 2003–2018 period is unsurprising given that by 2012, average research quality scores were closer across universities and disciplines than in 2003. One advantage of this approach is that it uses information about only AQSs for universities and disciplines.12 All that is needed is a measure of the research quality of the units. However, if individual data are available, it is possible to examine the separate contributions to the convergence process of exits, entrants and quality transformations. Using the decomposition described in Section 4, regressions of these contributions can be carried out separately, using the same basic specification. Differential rates of exit, entry and quality transformation can contribute towards convergence for several reasons. For example, there is an upper limit to the extent to which exits can continue to contribute to improving AQS: as the number of low-quality researchers is reduced, the scope to improve an AQS from exits diminishes. Similarly, regarding quality transformations, there may be diminishing benefits from initiatives that improve the research environment and skills of incumbent researchers. For entrants, there is a diminishing effect from a university's budget constraint, as well as from constraints imposed by other university requirements (such as teaching and administration staffing requirements).13 7 | ASSESSING SUSTAINABILITY OF RESEARCHER DYNAMICS AND THE IMPACT OF PBRFS The analyses of flows have emphasised their close links with the dynamics which could be expected in response to incentives created by a PBRFS. This section shows how the observed dynamics can be used to generate an implied equilibrium stock of researchers, whereby the distribution of researchers across QCs remains constant. This equilibrium distribution can be compared with the actual stocks observed at the end of each assessment period. A large disparity and an unrealistic equilibrium distribution suggest that the changes cannot be sustained indefinitely. Hence, it is reasonable to argue that the changes are sufficiently large to have been, at least to some extent, induced by the new incentives.

Using the transition matrix and vector of births in Table 1 gives equilibrium stocks of A, B, C and R researchers, respectively, for each of the periods 2003–2012 and 2012–2018, shown in Table 4. The equilibrium total stocks for the 2003–2012 transitions exceed actual stocks in 2012 by about 23%. It is, therefore, reasonable to suggest that the transition dynamics after 2003 represents to some extent a structural shift in response to the PBRFS, and this path cannot be expected to continue indefinitely. The equilibrium total stocks for the 2012 to 2018 transitions exceed actual stocks in 2018 by about 26%. Although the nature of the dynamics changed during the second period, they continued to generate large differences between actual and equilibrium stocks. This is also illustrated by the differences between the actual AQSs in 2012 and 2018 and the implied equilibrium AQSs. The value of the implied equilibrium AQSs is 4.46 for the 2003–2012 dynamics and 3.90 for the 2012–2018 dynamics. The actual values are 4.55 for 2012 and 4.91 for 2018. Both values are in excess of the equilibrium values. The AQS-increasing strategy depends on the initial AQS, so that to continue with fixed transition proportions is not optimal. The same approach can of course be applied to individual universities and discipline groups. 8 | CONCLUSIONS This paper has presented several methods which can be used to evaluate the effects on university research quality of the introduction of a PBRFS. The starting point was the fundamental role of staff turnover, and researcher transformations, in generating improvements in research performance. This led to an emphasis on methods associated with a social accounting framework. Stress was placed on an evaluation which exploits the nature of the likely responses arising from the particular incentives created by the PBRFS. These are used to motivate and inform empirical analyses and tests and to inform policy. The analyses are directly applicable to PBRFSs which are based on the evaluation of individual researchers and rely on the use of longitudinal data, collected over a period containing two or more assessment rounds. However, even where some form of group assessment is used, the methods are of value because changes in group measured quality depend on the same kind of flows and quality transformations. The PBRFS metrics used to evaluate research performance and to allocate funding to universities create specific individual incentives and stimulate organisational and management changes. Hence, even where individuals are not explicitly assessed in the evaluation procedure, they are inevitably assessed by current and potential employing units in their recruitment and promotion procedures. Hence, it would be of value, in such group evaluation processes, to collect the kind of information relating to individuals that enables the methods used here to be applied. The paper has shown how changes in the quality measure of a research unit can be decomposed into separate contributions of entries, exits and transformations of individuals. Furthermore, quality changes can be decomposed into contributions arising from changes in the discipline composition within a university and changes in the research quality of discipline groups. The question of whether PBRFSs give rise to a process of convergence, or divergence, in research quality can be evaluated using a regression specification. The approach can allow for testing of differences between discipline groups and universities and can allow for different lag structures. In addition, it is possible to investigate the contributions to the process of convergence of the components of staff turnover and quality transformation of individuals. In addition to directly linking changes to the new incentives, the sustainability of changes was examined. If the changes are found to be unsustainable, there is prima facie evidence to suggest that they were caused by the introduction of the PBRFS. The methods used here enable a deeper understanding of institutional responses to PBRFSs, the sustainability of changes in research quality and the contributions of changes in researcher quality and discipline composition to changes in institutional performance. The approach, involving explicit consideration of the special incentives for structural change which face each type of research unit, can contribute to the debate about the appropriate design of a PBRFS, given the policy objectives involved. The methods, therefore, contribute both to a better understanding of the consequences of existing schemes and, by focussing explicitly on the nature of incentives, to the design of performance-based funding schemes more generally. The need for consistency between clearly stated objectives and the incentives created by a scheme is paramount. Importantly, the research methods discussed in this paper can be applied to any type of performance-based funding scheme, irrespective of the way individuals are assigned to quality categories.

Research can be conducted using either inductive or deductive approaches. While inductive research aims to develop a theory, deductive research focuses on testing an existing theory.(1) Much of biomedical science research is conducted using the deductive approach through basic, applied, and in vitro or in vivo research and clinical trials. But what if the research aim was not to assess the relative efficacy and safety of new versus old treatments in a randomised controlled trial? What if the goal was, instead, to explore the reasons for rising tensions between the Emergency Department (ED) and General Internal Medicine (GIM) physicians in a particular hospital?(2) In such an instance, it would seem that methods such as randomisation and controlling for confounding factors might not be as appropriate for such a research question, as it deals with complex social phenomena rather than cells, tissues or pathogens. Instead, an inductive approach using qualitative research methods to explore processes, phenomena and settings would be more appropriate,(3) as healthcare systems are, after all, complex social organisations where power, inequality, conflict, competition and collaboration exist.(4,5) These realities do not lend themselves easily to deductive approaches measuring discrete variables, but they are important considerations, as social relations and interactions in healthcare organisations impact the outcome and cost of patient care.(3) Similar to how a randomised controlled trial might assess the efficacy of a new drug, qualitative research can explicate social phenomena within healthcare settings that impact patients’ quality of life, such as how palliative care teamwork might affect the quality of life of patients receiving palliative care.(6) Qualitative research findings may challenge existing perspectives and, therefore, could offer healthcare professionals (HCPs) and managers new and valuable insights. For example, through qualitative inquiry, Carter et al discovered that healthcare quality collaboratives are not all about collaboration – free riding and competition also abound.(7) Similarly, Knowles et al found that physical co-location of different HCPs does not automatically make them work together better or lead to better care.(8) This review focuses on a qualitative methodology called institutional ethnography (IE). It is meant for HCPs who are novice qualitative researchers interested in topics that include but are not limited to medical education and interprofessional collaboration. We cover what IE is, why it matters and how to apply it, using examples to illustrate key points. WHAT IS INSTITUTIONAL ETHNOGRAPHY? IE seeks to understand and capture, in detail, actual work processes performed by members of the organisation, and then trace how these work processes are coordinated at a higher level by policies, protocols, standards, competency frameworks and social norms. People may or may not be aware of these higherlevel influences on their work. Of interest to an IE researcher is the dissonance between what institutions think people ought to do and what individuals are actually doing on the ground. Being able to identify what has been lost when people try to translate policies and what could be done to rectify the situation is the unique value afforded by IE.(9-11) IE is not simply ethnography or focused ethnography(12) conducted in a social organisation such as a family, a school or a healthcare system. Similar to ethnography, IE is committed to ‘careful descriptive research’, a stance that is characteristic of ethnography. IE differs from ethnography in its goal. The overall goal of IE is to understand how the institutional arrangements of a society expressed in textual forms – such as government policies and organisational directives – affect the everyday work experiences of people.(13) This goal is clearly different from that of ethnography, whose goal is to delineate the cultural context in which a specific social behaviour takes place and makes sense.(14) For example, to understand the barriers to interprofessional collaboration in a healthcare practice, ethnographers examine what the term interprofessional collaboration means to different HCPs involved, how they define their roles in the collaboration vis-à-vis the roles of others, how they interact with others and how they interpret the actions of others. By contrast, institutional ethnographers start with a specific problem experienced by one group of HCPs in the course of collaboration with others (e.g. conflicting role expectations from other professionals and from the head of their own professional group). They then trace the causes of these problems to higher levels of institutional arrangements such as reporting structure and key performance indicators set for this group in the organisation that may impede effective interprofessional collaboration. As such, while ethnographic research starts with individual actors and ends its analysis at the individual level, institutional ethnographic research starts with individual actors but ends its analysis at the institutional level. Also, unlike ethnography, which does not have a unifying theme across studies of different cultures, each account of IE adds to a cumulative body of knowledge on how larger social institutions of a society enter into and shape the everyday work experiences of people in that society.(15) For example, if one IE study reveals reporting structure and key performance indicators set by the organisation as two significant institutional forces hindering interprofessional collaboration, while another IE study teases out the influences of more extended dimensions of institutional arrangements such as national healthcare policies, the knowledge generated by each study helps to render more visible the institutional forces governing individual experiences of interprofessional collaboration. IE’s focus on the everyday work experiences of people is a result of its theoretical affiliation with Marxist historical materialism, which argues that the most important social relationship is the relationship of production, which is work.(16) What then defines work? It is useful to take a step back and note that IE’s definition of work is broader than Marx’s original definition of the concept as wage labour. Besides paid employment, institutional ethnographers also consider unofficial/unpaid activities as work, so long as individuals take time and effort to engage in them.(11,15,17) Examples of IE-defined work include a mother attending school meetings to discuss her child’s access to disability support,(17) or the time patients spend waiting for test results at hospitals or clinics.(11) Given IE’s focus on everyday work processes, the research question does not arise from extant literature.(11) Instead, it comes from the dissonance between what a researcher observes to be happening in real life and what authoritative knowledge claims is happening or should be happening.(10,11) This dissonance is called a ‘disjuncture’ (Table I). When the disjuncture causes a problem from the ‘standpoints’ (i.e. social positions) of particular individuals, it is called a ‘problematic’, which is equivalent to a conventional research problem.(10,11) In essence, a problematic points to the social experiences that people encounter as troubling or difficult.(18) Once a problematic has been identified, IE researchers begin by iteratively collecting data describing the individuals’ standpoint, before broadening data collection to include their colleagues and remote collaborators.(10,11) Data collection methods include observations, interviews and focus group discussions (FGDs), along with identification of ‘local’ and ‘extralocal’ texts that coordinate the individuals’ work processes.(9-11) The term ‘texts’ refers to documents in spoken, written or graphic forms, such as policies, protocols, standards and competency frameworks.(19) Local texts refer to documents generated and used by individuals in their everyday work, for instance, healthcare staff writing notes about patients. Extra-local texts refer to documents disseminated from authorities, such as clinical guidelines set by the health ministry. Data analysis involves describing in detail the work processes of the individuals studied and tracing their everyday activities to extra-local texts to explicate the ruling relations that organise the work they do.(17) To ‘explicate’ is to describe the workings of a process that is hard to uncover, while the phrase ‘ruling relations’ refers to how ‘texts’ (i.e. documents) spell out how people are supposed to work together. AN EXAMPLE OF INSTITUTIONAL ETHNOGRAPHY One IE study sought to understand a problematic arising from intra-professional tension between hospital physicians at the ED and GIM departments in Ontario, Canada, in the early 2010s.(2) To understand the problematic – why the social relations were poor from the standpoint of the physicians – institutional ethnographers interviewed and shadowed the ED and GIM staff to understand their everyday work processes. They also identified and analysed local texts such as the physicians’ notes on patients.(2) IE researchers discovered that tension began when ED physicians started admitting many frail and elderly patients without concrete diagnoses. These patients were admitted from the ED for social reasons, as opposed to clearly defined medical reasons. These patients were admitted mostly to GIM wards, adding to the busy workload and exacerbating the existing shortage of beds in these wards. As there was no clear medical need to admit these patients, many GIM physicians felt that the extra patient load had wasted their time, and consequently, tension arose between them and their ED counterparts.(2) Following IE methods, researchers identified and explicated how an extra-local text governed the ruling relations of the work processes for ED and GIM physicians. This text was a new government policy mandating shorter wait times at EDs, which led to ED physicians feeling compelled to admit or discharge patients quickly, thus resulting in a change in their patient admission behaviours.(2) In the parlance of IE, this mandate, in the form of a text, coordinated the social relations between the ED and GIM physicians.(2) WHY DOES INSTITUTIONAL ETHNOGRAPHY MATTER? IE asserts that the work processes of different groups of people on the ground are coordinated by extra-local texts that individuals might not be fully aware of.(10,11) In the preceding example, the GIM physicians may or may not have been aware of the new government policy mandating shorter wait times in the ED. The GIM physicians only knew that they were experiencing a greater workload and felt resentment towards their ED counterparts for not making concrete diagnoses before admitting patients. The diminishing goodwill between the two groups of physicians thus formed the problematic of the IE study. It is also significant that the mandated shorter wait times in the ED did not address the underlying issue of patients being sent to hospitals for social rather than medical reasons;(2) the government policy only moved the workload from the ED to the GIM department. This disjuncture marks the dissonance between the official understanding of an improved hospital experience (shorter wait times) and the actual work being done on the ground (ED patients being moved to the GIM department without any problem resolution). As shown in this study,(2) through understanding people’s actual work processes on the ground, IE methods helped to uncover the link between the government mandate (extra-local text) and the resulting dysfunctional work processes affecting the ED and GIM departments which, in turn, led to heightened intra-professional tension between both departments. In short, IE serves as a practical qualitative research methodology that helps to trace everyday work processes to higher-level coordinators such as institutional leadership and management. Through this, individuals can become aware of their position in the larger systems and potentially have the opportunity to enact change and bring about new approaches to their work.(17) APPLICATION TO INTERPROFESSIONAL RESEARCH: A WORKED EXAMPLE IE is increasingly used in health services research(20) but has yet to penetrate the realm of healthcare professionals.(11) Given that healthcare professionals and scholars may be unfamiliar with IE and thus may find it challenging to employ, this section seeks to provide a worked example of how an IE study on interprofessional research was conducted. Study background The IE study to be analysed was conducted by Braaf et al.(21) They focused on the time-out procedure recommended by the World Health Organization (WHO)(22) to counter the rise in avoidable surgical complications(23) and adverse events(24-26) due to suboptimal communication among interdisciplinary surgical teams. In essence, the time-out procedure is a brief pause taken before a surgery begins, wherein the whole interdisciplinary surgical team comprising the surgeon, nurse and anaesthetist are supposed to check the patient’s identity, confirm the operative site and side by inspection, and ascertain the type of surgery to be performed.

Step 1: Identifying the disjuncture and problematic IE studies begin with the identification of a disjunction, that is, the gap between what is actually happening and what authoritative knowledge claims is or should be happening.(10,11) Braaf et al sought to understand why interdisciplinary surgical teams failed to adhere to the time-out checklist despite the fact that Australian hospitals had incorporated it into preoperative checklists as per the WHO’s recommendations.(21) This was an issue of research significance, because the disjuncture’s resulting problematic was that surgical errors continued to occur despite time-out being implemented. From the perspective of research question formulation, IE is suitable for busy clinicians who want to address a problem they have observed or experienced in the workplace – specifically, problems that result when things that are officially supposed to be done do not happen in practice. One example would be how (pre-licensure) interprofessional education does not always lead to collaborative practice in clinical settings.(27) IE would be useful to address such research problems, as there may be texts that could explicate ruling relations that do not encourage interprofessional teamwork.

Step 2: Data collection methods The methods for data collection used by Braaf et al for their IE study included observations, interviews and FGDs.(21) Braaf et al observed participants (interdisciplinary surgical HCPs) for 2–4 hours during mornings, afternoons, weekdays and weekends, totalling 350 hours.(21) These observations were made at a distance to enable the informant’s speech to be heard clearly, but not so close to be intrusive, cause disruption or contaminate sterile areas.(21) The observer noted the extent to which different HCPs adhered to the time-out procedure that was supposed to be implemented before each operation.(21) As a data collection method, observation is useful because it is common for individuals to say they are doing one thing when in reality they are doing something else, not necessarily because they are dishonest but because they may lack awareness or may be unable to articulate the subtleties of what goes on during their interactions with others.(28) Observations place researchers at the centre of the action, where they can see as well as hear what goes on.(28) Besides observations, Braaf et al also conducted interviews and facilitated FGDs. A total of 30 participants took part in the interviews or FGDs. Each interview or FGD lasted for about 40 minutes. Of these 30 participants, 12 also agreed to be observed for the purpose of data triangulation.(21) Braaf et al did not provide demographic details of the interview and FGD participants. However, in general, interviews are useful, as this method of data collection helps researchers gain insight into individuals’ experiences;(29) in this case, their workplace experiences could explicate ruling relations, that is, the identification of remotely crafted texts such as policies and protocols that determine the actual day-to-day work that people have to do. FGD also helps in obtaining a detailed understanding of processes,(29) in this instance, processes that govern participants’ workplace relations. FGDs differ from interviews in that their additional group dynamics and interactions among the participants(29) could presumably help researchers to appreciate the processes from a multiple-departmental perspective.

Step 3: Data analysis method Braaf et al then explicated the ruling relations that undermined the effectiveness of the time-out procedure. By triangulating the data collected through observations, interviews and FGDs, they found that although the interdisciplinary surgical team attempted to activate the text on quality and safety, their attempt was overwhelmed by more powerful competing extra-local texts on productivity and efficiency, specifically the state’s elective surgery access policy(21) and a document spelling out the key organisational performance indicators to be accomplished.(21) The researchers’ knowledge of hospital administrative processes also enabled them to triangulate regular audits (that ensure maximal utilisation of theatre time and space) with interview data, which showed that most surgeons and anaesthetists found the time-out procedure time-consuming. This led them to avoid performing it, although it took less than a minute to perform. Out of 107 surgeries observed, the entire time-out procedure was performed only 11 times. It was not performed for five surgeries, and in the remaining 91 surgeries, the procedure was either abridged (with omission of certain steps) or incomplete, where members of the team did not participate.(21)

Step 4: Recommendations and conclusions useful for changing practice Based on the findings of their IE study, Braaf et al made the following recommendations to improve the implementation of time-out procedures. First, they proposed that hospital leadership should implement communication education programmes that seek to flatten extant hierarchies and promote tolerance for open questioning by co-workers. Second, they suggested that government departments crafting healthcare policies that determine hospital performance indicators must take into account the communication challenges faced by surgical teams for delivering safe patient care in constrained timeframes.(21) IE is, thus, a useful problem-solving methodology to flag underlining issues. By capturing in detail how time-out was actually performed in busy operating theatre environments and tracing the extra-local texts that explicated the (hidden) ruling relations coordinating the work processes of interdisciplinary teams, IE allowed Braaf et al to generate concrete evidence to support their recommendations.(21) In the present era, where evidence-based approaches reign supreme, IE provides qualitative HCPs and researchers a useful tool to explicate relations and discuss change with key stakeholders. In this example, the findings of this IE study opened up new avenues for understanding and solving the time-out problem in operating theatres. Instead of following the official time-out guidelines and focusing on how to improve adherence to these guidelines, Braaf et al demonstrated the need for government agencies to consider the realities of interprofessional communication when defining key hospital performance indicators. Additionally, they showed that the deep-rooted power asymmetry and hierarchy between doctors and nurses in the hospitals was a factor that hindered the nurses from performing their expected role in successfully leading the time-out communication, even when time pressure was not an issue. These findings make it apparent that efforts to improve patient safety by highlighting the perceived importance of time-out among HCPs in the operating theatre are unlikely to be fruitful without dealing with power and hierarchy. The actionable insights revealed by IE render it more useful than other qualitative methodologies.

OTHER CHALLENGES

HCPs who intend to use IE may face some challenges. First, IE studies require significant resources and commitment to conduct, and the translation of findings also requires engagement with stakeholders such as hospital administration and policymakers.(30) Second, as healthcare professionals are trained to consider randomised controlled trials as the gold standard for scientific research, IE studies, which have comparatively much smaller sample sizes, may be deemed as having limited reliability and generalisability. Another perceived limitation is that in IE studies, the researcher may collect data from participants and even analyse it. This practice may lead researchers who are unfamiliar with qualitative methodologies to think that the high degree of subjectivity and bias will render the findings invalid. However, as pointed out by Cristancho et al, qualitative research has its own set of quality criteria to ensure trustworthiness and rigor.(3) Adapting Frambach et al’s(31) framework, we illustrate in Table II how IE studies, using the study by Braaf et al as an example, fulfil quality criteria.

CONCLUSION IE is a useful qualitative methodology that enables healthcare professionals and researchers to trace everyday work processes to higher-level coordinators (such as institutional leadership and management) that individuals may or may not be aware of. Through such means, stakeholders could then adopt concrete, actionable improvements that will benefit institutions as well as individual healthcare professionals.

The World-Wide Web provides opportunities for users to become publishers and distributors of their own creations(Van Dijck, 2009). A great deal of the information on the Web is content created by its own users, being at the same timesources and receivers (Morris & Ogan, 1996). With the development of Web 2.0, social networking sites appeared and soonbecame the Web users’ favorite places to produce and share information, and to communicate with each other (Ortega,2016). Regular social networking sites such as Twitter and Facebook allow users to view a great range of content, with postsand interactions often of a personal nature shared amongst family and friends, and are generally not academically focused(Citrome, 2015).For the benefit of scholars, wide-ranging web based academic communities was born during the period 2006–2008(Ortega, 2016). As a more recent development, researchers have been increasingly adopting academic social networking sites in their professional lives for research-related activities (Gruzd, Staves, & Wilk, 2012). Successful adoption of thesesites by individual academics seems to reflect a combination of academic capital and social networking skill (Thelwall &Kousha, 2014). ResearchGate (RG) and Academia.edu are two well-known academic social networking sites, which combine communication and dissemination, by incorporating a repository for academics’ publications (Thelwall & Kousha, 2015b).According to Alexa.com, the popularity of both sites, which are calculated using a combination of average daily visitorsand page views, have shown a significant increase over the past 3 years (Nov 2013–Oct 2016) (Alexa.com, 2016; Thelwall& Kousha, 2015b). RG has been the place where the world’s scientists and researchers come to solve research problems connect with their peers, and to seek collaboration.RG has verified more than 12+ million scientist or research professional members (ResearchGate, 2017). It provides aplatform for academic users to share publications and knowledge, to connect and collaborate with colleagues, to get readingand citation stats for their works, to ask questions and get answers for research problems, to search and monitor peers, andto even find suitable academic jobs (Giglia, 2011; Muscanell & Utz, 2015). Studies have revealed the main usage reasons activities and academic discipline characteristics of RG users (Chakraborty, 2012; Elsayed, 2016; Nicholas et al., 2015; Ortega,2015b; Stachowiak, 2014; Van Noorden, 2014), and examined the effectiveness of RG metrics for individual, institutional land national research rank (Martín-Martín et al., 2016; Onyancha, 2015; Orduna-Malea, Martín-Martín, Thelwall, & DelgadoLópez-Cózar, 2017; Thelwall & Kousha, 2015b; Yu, Wu, Alhalabi, Kao, & Wu, 2016), but these results were calculated withinsmall or raw samples, which included zombie users (who only registered their accounts actively or passively, but who donot actually participate in or utilize RG). A greater focus on the unique characteristics and utilities of specific academic social networking sites and a more robust understanding of scholars’ use preferences and practices is warranted in future andongoing research (Williams & Woodacre, 2016). Given the debate of the reasonableness of utilizing RG metrics for indicating individual reputation (Muscanell & Utz, 2015; Orduna-Malea et al., 2017; Yu et al., 2016), discussion of whether RG usagemirrors the actual research activity level of institutions still remains sparse Based on data collection and filtration of RG users from different U.S. research universities, this study is important inthat it focuses on the institutional dimension, revealing user participation and behavior differences in RG reputationalmetrics among universities at different levels of research activity. It reduces the interference of individual outliers, and takes the university members as a whole to examine the current structure of active RG users, as well as the effectiveness of RG metrics for institutional research activity level evaluation. In addition, it explores the institutional academic socialnetwork amongst these institutions. Such understanding is helpful for a deeper discussion of RG metrics, meaningful userparticipation, extending academic interaction, and improving user experience on academic social networking sites.

2. Literature review

2.1. Academic social networking sites

Because of collaborative trends and the explosive growth in the number of publications, academic communication isundergoing a transformation from journals and conference presentations to web interaction (Liu, 2003; Mas-Bleda, Thelwall,Kousha, & Aguillo, 2014). As a kind of social networking service (SNS), academic social networking sites can not only be used for communication, and maintaining and developing professional relationships, but also for listing, storing and sharingindividuals’ publications, which highlight their academic characteristics (Bianchini, 2012; Mas-Bleda, Thelwall, Kousha, &Aguillo, 2013; Thelwall & Kousha, 2015b). They allow a certain responsiveness and informality that is not possible with theformal publishing process (Ovadia, 2014).Academic social networking sites were originally extensions of personal homepages that incorporated social networkingfeatures and allowed academics to list and share, such as Mendeley and Connotea (Henning & Reichelt, 2008; Hull, Pettifer, &Kell, 2008). With the development of network technology and services, other sites, such as Academia.edu and RG, appear to beprimarily spaces for academics to describe themselves, to promote their works and reputations, to make new connections and to collaborate with peers (Gruzd et al., 2012; Campos-Freire and Rúas-Araújo, 2016; Thelwall & Kousha, 2015b). Additionally,a wide range of social media metrics (such as reads, downloads, skills, etc.) for scholars and outputs related to scholarlyactivities, are utilized by academic social networking sites (Haunschild & Bornmann, 2016; Thelwall & Kousha,2015). Animportant metric is citations, and RG finds more citations for recent articles than both Web of Science and Scopus, althoughless than Google scholar (Thelwall & Kousha, 2017).Research on academic social networking sites mainly falls into two primary arenas – promises (i.e. potential benefits to theacademic community) and perils (i.e. reservations expressed by scholars) (Williams & Woodacre, 2016). As for the promises,academic social networking sites encourage sharing and extending research to the public (Morrison, 2010), promote partici-pation and collaboration (Mangan, 2012; Veletsianos, 2013), create scholarly identities (Gruzd & Goertzen, 2013), and makepersonalized recommendations (Kincaid, 2011). For the perils, there is information overload (Gruzd & Goertzen, 2013), andprivacy issues (Veletsianos & Kimmons, 2013). In addition, the meaning of professional discussions (Friesen & Lowe, 2012;Tess, 2013), and the quality of contributions and publications (Mangan, 2012) to academic social networking sites are oftendoubted and questioned. Research on the users of academic social networking sites mainly focuses on the participation and behavior differences between different demographics and disciplines. More senior researchers were found less likely to join academic socialnetworking sites such as Academia.edu and Mendeley. However, scientists having one kind of social web profile were more likely to have another in many cases, especially in the life sciences and engineering. (Mas-Bleda et al., 2014). Thelwall &Kousha (2014) indicated that faculty members tend to attract more profile views in comparison to students, but femalephilosophers did not attract more profile views than did males. Ortega (2015a) further revealed that disciplinary differences of users are observed across every platform.After reviewing the descriptions, primary purposes and use capabilities, and strengths and weaknesses of the popular sites(including Academia.edu, Mendeley.com, ResearchGate.net, Zotero.org and Google Scholar), Williams & Woodacre (2016)indicated that the majority of available articles broadly overview the adoption of social media by scholarly communities, but few focused on specific platforms, and even fewer offered practical advice or guidelines for use within higher education. Hence, deeper studies about university members’ participation and characteristics on specific academic social networkingsites are needed and helpful for revealing the status quo, and for exploring the use of academic social network sites for reputation promotion.

2.2. Specific academic social networking site – RG

Compared to other academic social networking sites, RG is the preferred platform for creating author profiles, and for archiving and sharing publications. In addition, RG is used more than Academia.edu in all areas except the Arts & Humanities(Bosman & Kramer, 2016).The increasing utilization of RG worldwide, and its well-designed network functions and services, encourage researchers to choose RG for case studies of academic social networking sites. Studies on RG can be mainly divided into two categories, surveys on RG, and the utilization of RG metrics.A worldwide survey conducted by Nature revealed that more than 88% of scientists and engineers were aware of RG (VanNoorden, 2014). However, a case study of faculty participation on RG showed that much fewer researchers had registered(Stachowiak, 2014). Although RG is considered as a tool to communicate research results and knowledge, the level of knowl-edge utilization as obtained from RG is still low (Corvello, Genovese, & Verteramo, 2014). The results of surveys revealed that the main reasons for using RG are obtaining and sharing articles, forming study groups, keeping up-to-date, and catching up with others’ research fields (Chakraborty, 2012; Elsayed, 2016; Nicholas et al., 2015). The main activities on RG are simply maintaining a profile, sending/reading messages, and participating in discussions (Chakraborty, 2012; Van Noorden, 2014).With the diversified presentation functions and user-centric services, RG is used more as a Facebook-like networking toolthan a Twitter-like communication tool (Hoffmann, Lutz, & Meckel, 2016). In addition, the results also revealed that academicdiscipline appears to play a role in defining researchers within RG (Elsayed, 2016; Ortega, 2015b; Stachowiak, 2014).Since the metrics on academic social networking sites have becoming newly effective indicators for academic impacts, leading to the emergence of the field of altmetrics (Ortega, 2015b; Priem, Taraborelli, Groth, & Neylon, 2010), studies alsohave focused on the effectiveness and utilization of the indicators provided by RG. RG provides the number of full-textdownloads, views and citations (based on information in its database) for registered publications; and also provides some information on individual members, such as the total number of publication views and downloads, as well as how many followers and followees they have (Kadriu, 2013). Compared to the other academic social networking sites, RG indicators were highly correlated to Mendeley readers, and Google Scholar Citations (Martín-Martín et al., 2016). In terms of papercoverage, university ranking and research impact, RG metrics showed high correlation with most of the widely adopted ones, such as those in Web of Science (Onyancha, 2015; Thelwall & Kousha, 2015b).Specifically, the RG score, which integrates both bibliometrics and altmetrics, by measuring researcher publications, questions asked and answered, and the number of followers, shows its potential to be an effective indicator for measuring both institutional and individual performance (Yu et al., 2016). The RG score is implicitly accepted because it is highly correlated with Google Scholar indices, and offers the possibility of building a wider media-based reputation (Muscanell &Utz, 2015; Nicholas et al., 2015). Using altmetrics by adopting a relational networking perspective, Hoffmann et al. (2016)found that junior faculty members more actively try to generate social capital through their communication and networkingin an effort to establish themselves, while senior academics tend to limit their engagement in listing their publications and other activities. This finding adds richness and differentiation to scientific impact assessment. However, some researchers questioned the RG score as a measure of scholarly reputation. Kraker and Lex (2015) and Jordan (2015) suggested that RG score is non-transparent and irreproducible, and the score incorporates the journal impactfactor to evaluate individual researchers. Orduna-Malea, Martín-Martín, and Delgado López-Cózar (2016), Orduna-Maleaet al. (2017) argued that RG scores also considered activities related to asking and answering questions on the site, and should not be mistaken for academic reputation indicators. Due to the undisclosed algorithm of the RG score and its potential update,the effectiveness of RG score as an academic reputation indicator still needs to be explored from multiple perspectives. To summarize, despite the fact that most research studies on RG have focused on a user survey of participation and site activities, very few revealed the institutional user structures and their behavior characteristics. The research methodsmainly included use of questionnaire surveys (Corvello et al., 2014; Elsayed, 2016) or open-ended interviews (Nicholaset al., 2015). Some utilized data from the RG site, but were based on a small sample or lacked data filtration (Hoffmann et al.,2016; Thelwall & Kousha, 2015b; Yu et al., 2016). Collectively, previous studies have shown the value and correlations of RGindicators to other accepted impact indicators, but there is not consensus on the most appropriate way to measure academicinfluence via social media (Jordan, 2015). Previous studies have also shown a gap in the research of user characteristics and behavior analysis, especially in revealing institutional differences and the social networks of institutions at different researchactivity levels.

3. Purpose and research questions

This study is to address the general question: given the popularity of research-oriented social networking sites in thescholarly community, does scholarly reputation on the RG social networking site realistically mirror the reputation ofresearch universities? It aims to verify the research attributes of RG, to identify the institutional differences between selectU.S. research universities, and to reveal the social network among the institutions in different research activity levels. Thespecific research questions include the following:

RQ1. To what extent do RG metrics reflect the institutional differences and the research activity level of a university?

RQ2. How does RG reflect the institutional academic social network of U.S. research universities?

4. Methods

RG uses a model whereby users can follow the network activity of other users, seeing things like papers added, questions answered and asked, and allowing user endorsements for specific skills (Pérez-Rosés & Sebé, 2015). To show the user characteristics and impact, RG provides a series of indicators on user profile pages, including publications, reads, citations, profile views, skills and expertise, followers and following (i.e. followees). In addition, RG grants a score based on the publications in a user’s profile and how other researchers interact with the user’s content (Citrome, 2015). RG claims thatits score measures scientific reputation based on an individual researcher’s contributions, interactions, and reputation (Yuet al., 2016). Hence, this study chose user demographic data (name, university) and the metrics previously mentioned as the measurements of each user, and utilized “train crawler” (locoy.com) to crawl the user data from the user’s profile page. Forexample, Table 1 is part of the obtained user data for members of “Brown University”. It includes the name, university, RGscore, publications, reads, citations, profile views, followers and followees of each user. The data collection and processing process is illustrated in Fig. 1.As shown in Fig. 1, this study chose U.S. research universities according to the Carnegie Classification of Institutions of Higher Education (2016). The Carnegie Classification is a framework used to identify and recognize institutional differences of U.S. universities and colleges in supporting programs and research, including classifying research universities into threelevels: highest research activity (R1), higher research activity (R2), and moderate research activity (R3). Given that the Carnegie system assigns research activity levels without a ranking, we used U.S. News National University Ranking (U.S.News, 2016) to help narrow down the study sample. Then the top 20 universities of each level were selected according to the national university rankings, which are comprehensively measured by indicators of academic quality as well asindicators such as graduation and retention rates, peer assessments, faculty and financial resources, and selectivity. Due toties, 61 universities were selected as the final sample of universities: 21 for R1, 20 for R2, and 20 for R3. The RG user datacollection procedure was divided into 2 steps:

•The first step aimed at obtaining the user’s RG data to answer RQ1. The crawler obtained URLs of user profiles from themembers page of each university, and then retrieved the demographics and RG metrics from the personal profile pageof each user. The crawling took place 22nd September–2nd October, 2016, and captured 168,059 users’ data from 61universities. Since RG score is calculated based on the personal research profile and interactions with other researchers(ResearchGate, 2017), a score below 0.01 can be regarded as an automatically generated score, or as an inactive “zombie”user. Hence, this study additionally filtered users whose RG scores were below 0.01 to only include the remaining 87,083(51.82%) users who were relatively active.

•The second step aimed at obtaining the institutional academic social network data and to answer RQ2. The crawler obtainedthe followers and followees data from the personal profile page of each user selected in the first step. Due to RG limitationsin presenting connectivity data on user profile pages, crawling followers and followees data was limited to what wasincluded in each user’s main profile page. Specifically, for every user in the sample, his/her top 10 followers and top 10 fol-lowees’ institution data were collected and combined according to institution. The crawling took place 21 st October–15thNovember, 2016, obtaining data for 459,763 followers and 360,250 followees. This study utilized Gephi 0.9.1 to do thesocial network analysis and visualization of the follower-followee relationship data.

5. Results

The data was analyzed and summarized in relation to RG reputational metrics: (a) RG score, (b) interaction metrics suchas profile views, number of followers, and number of followees (i.e., number of following actions), and (c) publication metricssuch as number of publications, reads, and citations. In addition, the academic social network based on the follower-follower relationships on RG is also obtained and analyzed.

5.1. RG score

In general, the institutional number of users, ratio of users with RG score higher than 0.01, and user RG scores decreased with the research activity level of the U.S. research universities. As shown in Table 3, both the median and mean of users with RG scores higher than 0.01 from R1 universities were much higher than those from R2 and R3. The ratio of users with RGscores higher than 0.01 also decreased with the research activity level from 56.78% (R1) to 26.94% (R3). As shown in Table 4,a similar pattern is observed for user RG scores among the institutions at different levels. The 95% confidence interval forthe mean value of RG score of R1 is between 17.38 and 18.52, R2 is between 13.85 and 15.30, and R3 is between 10.75 and12.70.Kruskal-Wallis H test is a non-parametric technique (distribution-free test) that can be used for both continuous and ordinal-level dependent variables (Pallant, 2005; StatisticsSolutions, 2017). In this study, Kruskal-Wallis H test and a Dunn-Bonferroni test for post hoc comparisons were utilized to determine if there were statistically significant differences in theRG scores of different universities. The results in Table 4 and Fig. 2 show that the differences in RG scores among differentresearch activity levels were relatively large (2= 44.723, df = 2, ES = 0.745). In addition, the fluctuations of the scores within each category illustrate institutional differences, indicating that an institution’s overall RG score does not coincide with itsU.S. News ranking within the same research level as the ranking is not entirely research-based.

5.2. Interaction metrics

Interaction metrics include profile views; number of followers, which reveals a user’s reputation and academic influence;number of followees, indicating a user’s motivation for information acquisition and interaction; and difference between thenumber of followers and the number of followees.Table 5 shows that both profile views and number of followers declined from R1 to R3, indicating users from higherresearch activity level universities received more attention on the social networking site. All the means of interaction metricsfall within their respective 95% confidence intervals. The Kruskal-Wallis H test results also indicate that the between groupdifferences in the number of followees (2= 3.455, df = 2, ES = 0.058) were less obvious than the other interaction metrics. R2universities had the highest median and mean in the number of followees per user. In addition, the standard deviations of allinteraction indicators of R3 were much higher than those of both R1 and R2, which explains more variations of interactionmetrics could be found among users from R3 universities.In addition, regardless of research activity level, the average number of followers per user was larger than that of followees(see Table 5 and Fig. 3). The difference between the numbers of followers and followees decreased with research activitylevel. This result suggests users from lower research activity level universities tend to acquire information by followingmore than those from higher level universities. Although R3 universities had obviously disparate distribution in Fig. 3, thedistributions of R1 and R2 were all relatively centralized. The reason for this result could be the relatively small user sampleof R3 universities compared to their counterparts in R1 and R2.

5.3. Publication metrics

Publication metrics include the number of publications that a user uploaded onto RG, indicating a user’s participationstatus and academic output; and the total numbers of reads and citations of these publications, revealing academic quality andimpact. As shown in Table 6, the number of publications per user and citations per publication increase with research activitylevel, while the number of reads per publication exhibit the opposite trend. All the means of publications metrics fall withintheir respective 95% confidence intervals. The between group differences in publications (2= 42.931, df = 2, ES = 0.716) and citations (2= 32.274, df = 2, ES = 0.538) were larger than reads (2= 1.297, df = 2, ES = 0.022). This result indicates thatusers from higher level universities tend to have more publications and academic influence. RG reads, however, has smallbetween group differences, and may be affected by many factors, such as a higher relative number of publications, recencyof publications, uploads of publications, and the accessibility of full-text.In addition, when examining the relationship between reads and citations, universities can mainly fall into 2 groups (Aand B) as shown in Fig. 4. Both group A and B had similar reads per publication (<60), while the difference was citationsper publication (A > 16, B < 16). The higher citations indicate academic quality, while the lower reads could be due to readsoutside RG for articles published earlier and not reflected in RG metrics. Also, active uploads of new publications and theunavailability of full-text could also contribute to low reads. The only outlier was Texas Wesleyan University (R302), whichhad far more reads per publication than others. It was attributed to some individuals’ publications which got extremely highreads, while the total member of the university was limited.The result showed that nearly all R1 universities (except University of Notre Dame, R116) were in Group A, while mostR2 and R3 universities were in group B. It demonstrates that the quality and academic influence of the publications of R1are relatively higher, while their publications update might also be frequent.

5.4. Academic social network formed among research universities in RG

The institutional follower-followee relationships data was obtained from the affiliated members of the 61 universi-ties. The academic social network analysis aimed to show the main institutional relations among universities in differentresearch activity levels. Since the data indicated that the self-follow relationship of institutions occupied a relatively highproportion of the follower-followee relationships for every university, this study excluded these relations to better showthe follower-followee relationships among institutions. In addition, two measures were taken after alternate names of thesame institutions were unified and combined to better display the institutional social network on RG. The first measurewas removing the relations with a weight of 1, which could be random or casual. The second measure was choosing andfocusing on the top 30 (sorted by weight and alphabetical order) followers and followees institutions of each university inthe sample. As a result, 513 institutions (nodes) and 3237 follower-followee relationships (edges) were obtained for socialnetwork analysis. In addition to the universities of R1, R2 and R3, some non-educational institutions and institutions outsideU.S. are also included, which formed the “Others” category.Fig. 5 is the social network obtained in Gephi using “Hu yifan” layout, which combines a force-directed model with agraph coarsening technique to reduce the complexity, and provides good quality for large graphs (Hu, 2005). The node sizeis determined by the weighted degree, which is the sum of the weighted in-degree (number of followers) and out-degree(number of followees) and reflects the connection frequencies of one institution with others. The node colors represent the different research activity levels. More frequent follower-followee pairs are connected by a thicker directed edge, while thedirection indicates the behavior of following. Fig. 5 displays the follower-followee relationships among R1, R2, R3 universi-ties, and other institutions. It is obvious that the weighted degrees of some R1 universities are greater than those of R2 andR3 universities. “Stanford university” (65 in-degree, 46 out-degree) and “Harvard university” (66 in-degree, 51 out-degree)are the two largest nodes, which have the highest weighted degrees of 12,886 and 12,032, respectively. Meanwhile, “Har-vard University” has the highest closeness centrality (0.45) and betweenness centrality (10262.82), which reveals its highreachability and variety in a member’s network (Hoffmann et al., 2016; Leydesdorff, 2007).Table 7 summarizes the institutional social network properties of universities at different research activity levels. Theresults indicate that the number of nodes and edges decline significantly along with the research activity level. R1 universitiesdramatically comprise the highest density (0.067), average degree (14.37), average weighted degree (659.12), and the averageclustering coefficient (0.342), and the lowest diameter (4) of the network. It illustrates that the connections among R1universities are much closer than R2 and R3, which can also be recognized in Fig. 5.Table 8 reveals the details about edges among different university research activity levels. It indicates that the interactionswith R1 universities are far greater than those with R2 and R3.To display the main relations among R1, R2 and R3 universities, this study further filtered out the “Others” nodes and thenodes which had a degree under 5. The force directed “ForceAtlas2” layout was implemented to allow a visual interpretationof the structure, and turn structural proximities into visual proximities, in Fig. 6. Nodes repulse each other like magnets,while edges attract their nodes, like springs. These forces create a movement that converges to a balanced state (Jacomy,Venturini, Heymann, & Bastian, 2014). It shows that U.S. universities are mainly distributed in three circles, R3 universitieswere in the outermost circle, R2 universities in the middle, while most of the R1 universities closely linked together in thecenter circle. The top 10 universities of each research activity level are labelled, and their network attributes are found inAppendix B. It is obvious that R1 universities have highest weighted degree and closeness centrality.

6. Discussion

Although the mission of RG is to connect the world of science and make research open to all (ResearchGate, 2017),whether RG is recognized and used as a research-oriented platform has not been systematically examined and confirmed.This study collected and compared RG metrics of users from U.S. research universities in three research activity levels (R1, R2 and R3) as defined in Carnegie Classification. The major findings are highlighted and discussed around the two researchquestions of this study.

6.1. RQ1. To what extent do RG metrics reflect the institutional differences and the research activity level of a university?

Among the three levels of universities, the total number of RG users, number of RG users with a RG score equal or greaterthan 0.01, and the proportion of the latter showed dramatic decline along with research activity level. This result supports theprevious finding that researchers in universities which have more influence in terms of research and web impact are moreactive in social media than those in universities with low research visibility and impact (Onyancha, 2015). It also confirmsthat researchers from higher level universities tend to embrace RG more, utilize it to disseminate and follow publications,and to get citations and improve scholarly reputation (Campos-Freire and Rúas-Araújo, 2016).As a comprehensive index for evaluating users’ overall scholarly reputation on RG, the decreasing trend in RG score amongR1, R2 and R3 universities suggests users from higher research activity level universities indeed demonstrate such reputation.Although there would be some discrepancies in indicating individual reputation (Kraker & Lex, 2015; Orduna-Malea et al.,2017), our results illustrate that the RG score can be effectively used as a reputation indicator for research universities. Thisstudy demonstrated that users from higher research activity level universities tend to be more consistently reflected onRG, while lower level ones exhibited more randomness in RG use. Some users, although not widespread, from lower leveluniversities, also find value in RG and have taken full advantage of it. Despite some randomness for users in R3 universities,the overall user participation and RG score distribution largely reflect the characteristics of different university researchactivity levels.In terms of interaction, the university research activity level positively ties to academic influence as reflected in profileviews and followers. Users from higher level universities attract more academic users for learning, interaction and solicitingpotential cooperation. However, the minor fluctuations in the average number of followees among R1, R2 and R3 indicatethat users from R2 universities are more likely to follow others, compared to users from higher level universities who tendto be informers, and users from lower level universities who tend to be inactive. It could also be explained that a user’sinformation acquisition and interaction motivation are prone to individual differences.It was noted in this study that the number of followers tend to be much larger than followees, suggesting for researchuniversity users, that RG is more inclined to be a representation platform rather than an interaction one, similar to thefindings of Hoffmann et al. (2016), who found that participants do not follow a large number of their peers on RG. This isespecially true for users from higher level universities, who are more likely to share than acquire on the site.RG metrics such as users’ profile views, followers, publications, and citations coincide with research activity levels. SinceRG publication metrics could be potentially useful for the assessment of articles or authors (Haustein et al., 2014; Thelwall & Kousha, 2015a), the result illustrates that users from higher level universities are more likely to share their high qualityresearch on RG and be valued by peers. This study found that RG reads is a complex metric that does not directly tie toresearch activity level. Such factors as fast updating and relatively higher number of new publications from higher researchactivity levels may contribute to low publication reads. Old documents read before RG was first launched, and therefore notreflected in the “reads” metric could also be a reason for unexpected limited reads. In addition, journal articles, conferencepapers, and other publications such as books, chapters, and technical reports are all regarded as publications on RG. AsCitrome (2015) emphasized, copyright concerns can make general uploading and downloading inadvisable. As a result, thefull-text uploading ratio of high quality publications could be limited (Nicholas et al., 2015), and more results prior to formalpublication could be read (Thelwall & Kousha, 2015a), which could lead to the number of reads per publication of lowerresearch activity level universities being higher.Adding to previous findings that rankings based on RG metrics correlated moderately well with other academic metricsand rankings of academic institutions (Thelwall & Kousha, 2015b), this study found within-group fluctuations in all RGmetrics among the universities which were ranked according to comprehensive strength in each category. It illustrates thatRG metrics are reasonably used as research-oriented academic influence indicators, similar to the findings of Yu et al. (2016),rather than comprehensive measurements for institutions. Further, this study found major between-group differences innearly all RG metrics. These results validate that there are institutional differences in overall participation, interaction andpublication metrics on RG. Specifically, the largest between-group differences were in RG score and publications, whiledifferences in the number of followees and reads per publication were relatively small.The results of this study not only confirm previous finding that RG metrics of users from a specific university could beused to indicate the esteem or influence within the academic social networking site (Thelwall and Kousha, 2015b), butalso suggest that RG metrics reflect and can be used to evaluate the intensity of a research activity level. It suggests thatthe traditional bibliometric analysis of the publications and citations of institutions (Russell & Rousseau, 2010) could beextended to social networking sites with RG metrics. For example, traditional citation data serves as the backbone of manyexisting institutional ranking systems (C¸ akır, Acartürk, Alas¸ ehir, & C¸ ilingir, 2015), and higher academic ranks are positivelyrelated to higher publication and citation scores (Amara, Landry, & Halilem, 2015). With RG metrics, such as RG score andprofile views, institutions can be more comprehensively evaluated, adding supporting evidence of reputation and academicinfluence.

6.2. RQ2. How does RG reflect the institutional academic social network of U.S. research universities?

With the social network analysis of the follower-followee relationships of RG users on the institutional level, differentlayers of interactions among different university research activity levels are demonstrated in the social network of U.S.research universities.The study found users of RG are following or followed by the users of the same universities most, consistent with thefact that participants primarily interact within their offline community, such as institutional colleagues, which also showslarge institutional homophily (Hoffmann et al., 2016). The simple reason could be that the social networks on RG are mainlyextended from the social network of real life, in accordance with other social media websites. It demonstrates that scientificcollaboration and interaction are still easier within a certain region, even on academic social networking sites.The study also found that the interactions among universities of higher levels of research activity are far greater than thoseof lower ones. It not only indicates again that users from higher level universities are more popular, but also suggests that akind of “academic club” of high research activity level has formed on RG. Specifically, users from R1 universities follow eachother more frequently and make closer connections among each other. Hence, R1 universities constitute the central circle ofthe institutional social network on RG. The central academic social network encourages R1 university users to acquire highquality knowledge resources and to seek high efficiency scientific peer collaborations. Users from R1 universities have takenfull advantage of RG to form academic social networks and additionally to utilize them to promote their academic researchand academic influence. On the other hand, users from R2 and R3 universities primarily follow users from R1 universities,which are indicated by the group connections of R2 → R1 (69.66%) and R3 → R1 (63.93%). It additionally verifies that keepingup-to-date and catching others’ research are main reasons for RG use (Elsayed, 2016; Nicholas et al., 2015).Although the sample size distributions of the 3 levels of universities would influence the network to some extent, thenetwork structure reflects the actual participation status of university members on RG. The findings coincide with the“Matthew effect” in science which reveals more visibility of contributions to science by scientists of acknowledged standing(Merton, 1968), and emphasizes the trends on institutional level. The results also suggest that an academic social networkwith those of high research activity level can be regarded as a credible academic resource. The expansion and promotion ofthe academic social network on RG can be considered a way to encourage research and collaboration.

6.3. Limitations

There are several limitations of this study. First, in this study, crawling followers and followees data was limited to whatwas included in each user’s main profile page with a maximum of 10 for followers and 10 for followees, due to limitationsinherent in RG. Second, although the study included all RG users affiliated with the institutions in the sample, it is unclearhow representative these users were to each institution because RG is an open and voluntary social networking site for researchers. Also, the large-scale sample of self-created profiles cannot be identity verified one by one. Third, this studyonly focused on the major user metrics, which may miss some specific perspectives of user performance on RG such astimeline of behaviors and contributions. Finally, due to the fact that the algorithm for the RG score still remains a mystery topublic, we cannot illustrate to what extent it reflects user’s reputation or participation level. In addition, although RG scoreis a continuous variable, it is not assigned on a linear scale. However, this study took RG score as an overall performanceindicator and still used the mean value of RG score to compare the differences of universities.

7. Conclusion and future research

This study reveals that as a research-oriented social networking site, RG realistically mirrors the scholarly reputationof research universities in the U.S. It further sheds light on the institutional differences between different research activitylevels. Additionally, this study reveals the social network formed among the U.S. research institutions on the site.The results show that users from higher level universities are taking full advantage of RG to promote reputation andencourage academic interactions and collaborations, while those from lower research level universities demonstrate lessactive participation and more varied activities on RG overall. The academic social network formed on RG echoes real life.Besides following peers from the same universities, RG users primarily follow others from higher levels for academic con-nections and keeping up-to-date. R1 universities constitute the central core of the academic social network, formulating an“academic club” of high research activity level.Based on the institutional differences and academic social network revealed in this study, RG metrics and a RG aca-demic social network prove to be effectively utilized as institutional evaluation tools for discerning research activity level.Embracing RG or other research-oriented academic social networking sites could positively promote academic influenceand acquisition of research resources.Future studies could take a closer look at the differences between universities within the same or similar researchactivity level, explore disciplinary differences on research-oriented social networking sites and other behavior analysis onspecific groups of RG users. Also, a nation-to-nation comparison could be conducted to find cultural differences on RG andother academic social networking sites. Finally, future research may examine how research institutions, other than researchuniversities, use academic social networking sites and how their uses may vary from those by research universities. A morecomprehensive and deeper social network analysis of all research institutions could also be explored.

Author’s contribution

Weiwei Yan: Conceived and designed the analysis; Collected the data; Performed the analysis; Wrote the paper.Yin Zhang: Conceived and designed the analysis; Wrote the paper.

The Declaration of Helsinki uniformly requires that all biomedical research involving human participants, includingresearch on identifiable human material or data, should be approved by an ethical review committee (1). Having evolved out of the scandalous unethical research practices of the midand late twentieth century, the ethics review of study protocols, by independent ethics committees (IECs) or institutional review boards (IRBs), has become the international standard of ethically and scientifically acceptable biomedical research. Today, concerns over the quality of the IRB function are increasing worldwide. Globalization of the clinical trial in the last decade has caused more people in developing countries to participate in trials, international efforts to assure participants’rights and safety. Accordingly, recently more guidelines and regulations are being produced and more organizing conferences on ethics reviews have been held. Nevertheless, voices of concern about the protection of research participants have been increasing among researchers in developing countries and among the overseeing bodies in developed countries. Most importantly, the IRB function, to ensure the quality of research in terms of both its scientific and ethical aspects, differs from place to place for various reasons. In spite of the growing importance of ethics review on research involving human participation, concrete information on the performance of IRBs is not available, even in the developed countries. This scarcity of data on the quality of IRB performance implies an unsatisfactory situation with respect to the overseeing system for human participant research. In Korea, the enactment of the KGCP (Korean Good Clinical Practice) in 1995 required that clinical trials be reviewed and continuously monitored by IRBs. Several studies have shown that the enforcement of the KGCP has exerted remark

ably positive influences on the infrastructure and quality of clinical trials in Korea (2-4). In January 2001, the Korean government revised the KGCP based upon the ICH-GCP(E6), an international ethics guideline for clinical trials (5). This revision required that legal and institutional bases be established to ensure that the constitution and operation of IRBs be standardized and upgraded to international levels (6). However, given their short history the current situation of IRBs in Korea is neither desirable nor satisfactory. Most IRBs confine their roles to review only drug-related research, leaving large area of human participant research untouched by any ethics review. Despite any standardizing effect of the KGCP regulation, actual IRB practice varies greatly from one institution to another, depending on each institution’s local situation, experience, and resources. While globalization and the expansion of clinical trials and biomedical research call for transparent and competent ethics review, no regulations or guidelines are available for ethics reviews of biomedical researches, except the KGCP for clinical trials in Korea. Set against this backdrop, in March 2002, IRB members in major hospitals, biomedical researchers, medical directors of pharmaceutical companies and officers from health authorities founded the Korean Association of Institutional Review Boards (KAIRB) under the auspice of the Korean Academy of Medical Sciences. The main mission of the KAIRB is to help Korean IRBs build up ethical review capacity to the international level. For this purpose, KAIRB adopted various initiatives, namely, the co-hosting of the Korea-NIH Conference on Ethical and Regulatory Aspects of Human Participant Research in June 2002, with the aim of developing strategies to improve IRB ethics review quality based on concrete data. Accordingly, KAIRB conducted the first nationwide survey to evaluate the current situation with respect to the structures of the IRBs and the actual review processes conducted.

METHODS

Based upon the data on IRBs submitted to the Health and Welfare Committee of the Korean Congress in 2001, 74 IRBs were chosen as correspondents of this nationwide survey. After developing a questionnaire containing 67 question items on constitution, review process, and institutional policy with respect to research participant protection, the questionnaire was sent to a nominated individual at each IRB; these were identified by telephone calling each IRB and requesting the name of the most knowledgeable person about the IRB. Conducted in April 2002 as the first project of the KAIRB, this survey explicitly notified individuals that the purpose of the survey was to improve the quality of ethics review, and that all information provided would be kept confidential, to enhance the reliability of the data obtained. Of the 74 IRBs mailed, 63 returned the questionnaire in due time (response rate: 63/74=85.1%). After the returned questionnaires had been reviewed for the completeness and correctness of information, a computerized database was constructed for statistical analysis. The data are presented as proportions in the Tables and figures because most of the data were categorical variables.

RESULTS

All returned questionnaires were found to be valid for the analysis. Accordingly, 63 questionnaires were subjected to the computer analysis. Establishment of the IRBs The survey demonstrated that most of the IRBs in Korea were established after the implementation of the KGCP in 1995. Of 63, only 14 IRBs (22%) existed before the enforcement of the KGCP. Thirty-five IRBs (56%) were established between 1995-1997, and 11 (17%) between 1998-2001. In most IRBs (90%), the president of the hospital is responsible for appointing members. In more than two thirds of IRBs, the chairperson of the IRB is appointed without an election process within the IRB: the presidents appoint the chair in 29 IRBs (46.0%) and vice-presidents automatically take the chair in 19 IRBs (30.2%). While 37 (58.7%) of IRBs have their own administrators charged to ensure efficient IRB operation, 24 (38.1%) have no administrative support from the institution. Constitution of IRB membership

Since the sound constitution of the IRB members is a prerequisite for proper review, the authors tried to collect detailed information on the composition of each IRB’s membership. The average number of members in an IRB in Korea is 12.6 (range: 7-30). The average percentage of members from medical fields is 79.4%, comprised as follows: physicians (58.5%; range 3-16), medical scientists (9.2%; range 0-4), pharmacists (8.3%; range 0-3), and nurses (3.2%; range 0-2). The average percentage of professionals for ethics review is 9.7% including religious professionals (8.3%; range 0-3) and lawyers (1.4%; range 0-2). Non-affiliated layperson members comprised 4.4% (range 0-5) (Table 1). The gender ratio of the members is 78/ 22 (M/F). While 90% of IRBs have professionals for ethics review, including religious and legal professionals, only 43% have layperson members. About 40% do not have members who are not affiliated to the institution. Education of IRB members and investigators

Of the 63 IRBs, only 11 provide education to their members, and only 7 provide education for investigators on a regular basis, at least once per year (Table 2). Ninety percent of IRBs replied that they have written Standard Operating Procedures (SOPs) for IRB review, which means that 10% breach the KGCP.

Regularity and frequency of IRB meetings

Twenty-eight IRBs (44%) hold meetings regularly, while 54% replied that they have irregular meetings (Table 3). Only 19 IRBs (30%) hold meetings once a month or more. These data mean that more than half of all IRBs are relatively inactive.

Scope of IRB review and review burden in 2001

To the question‘, What kinds of research should be reviewed in each IRB?’, about 30% of IRBs make it a rule to review studies on epidemiology or on genetics and 33% of IRBs review studies involving stored biological samples. Only 21% review behavioral studies. On the other hand, 84% of IRBs replied that they should review drug studies, including postmarketing surveillance, 79% review new drug studies, and 62% medical device studies (Fig. 1). When asked about the number of research protocols reviewed in 2001, 24 IRBs (38%) replied that they reviewed 1-5 protocols, 11 (17.5%) reviewed 6-10 protocols, and 8 (12.7%) reviewed 11-20 protocols (Table 4). Nine IRBs reviewed more than 50 protocols in 2001. Academic research was reviewed at 19 IRBs (30.2%), of which only 2 (3.2%) dealt with more than 10 academic research protocols in 2001.

The review process

To the question,‘ Whether and how the protocols are reviewed before a formal meeting?’, about 41.3% of IRBs replied that all the members attend a formal meeting after reading

all protocols. Fourteen percent of IRBs adopt a primary review system, in which the protocols are thoroughly reviewed before the formal meeting. In about 30% of IRBs, the members do not read the protocols before the meeting; some of those IRBs replace reading with an investigator’s presentation (Table 5). Layperson members participate actively in the meeting in 54% of IRBs, and moderately understand protocols. In 54 IRBs (86%), principal investigators are allowed to attend IRB meetings only when asked by the IRB (52.4%), while they are not allowed to attend in 6 (9.5%) IRBs. If an IRB member is the investigator of a protocol under review, the member cannot attend the meeting in 9 IRBs (14.3%), or the member may attend to reply to questions but cannot review in 53 IRBs (80.9%) (Table 6). Only one IRB allows a member, who is also the investigator, to attend and review the protocol. The IRB decision is reached by consensus in 36 IRBs (57%), and by a majority vote in 24 (38%) (Table 7).

Continuing review

Thirty-four IRBs (54%) perform continuing review, while 26 IRBs (41%) do not. The patterns of continuing review among the IRBs are quite different. Some IRBs have reasonable criteria for the review interval for each protocol, but others have fixed review intervals for all protocols. The interval of the continuing review varies from every three months to once per year (Fig. 2).

Expedite review

Most IRBs in Korea have an expedite review system to ensure that appropriate action is taken after initial review of research protocols. Of 56 IRBs (89%) that have an expedite review system, 27 IRBs (43%) use a subcommittee system or pre-assigned members to conduct expedite review, and 19 (30%) collect opinions from individual members. Immediately reported serious adverse drug reaction is the most important topic category for expedite review (81%), but the categorizations used at expedite reviews vary among institutions (Table 8).

Policy for protecting human participants from research risk

To the question concerning the review of compensation for research participants, 46 IRBs (73%) review only mone- tary payment, and 7 IRBs (11%) do not review any kinds of compensation (Table 9). Only 22 IRBs (35%) make it obligatory for sponsors to have insurance for indemnity against participant injury. In terms of academic research, 35 institutions (62%) do not have any indemnifying policy for research participants (Fig. 3). Since non-response rates were high for questions on payment policy, it is difficult to form an overall picture on the current payment situation. Institutions have various payment policies, which vary according to the phase of a clinical trial (Fig. 4). For those who have not completed study, 9-19% of institutions are allowing for non-payment for the participants.

Self-reported problems of each IRB

To the open question asking about the problems and difficulties that each IRB faces, respondents replies ranged from one to several items (Fig. 5).

DISCUSSION

This study reveals many shortcomings in the Korean IRB system. The problems found by this survey can be summarized into three categories: (1) the structure of the ethics review system for research involving human subjects; (2) the review process; and (3) IRB policy with respect to protection of research participants. Problems in the structure of ethics review system for research involving human participants The most serious problem is limited scope of the IRBs review. All human-related research is not reviewed by IRBs; a very few categories of human research are covered by the IRB review system. Most IRBs limited their review only to legally bound research. Researches seeking Korean Food and Drug Administration (KFDA) approval for clinical trials upon new drugs, biologics, or devices must undergo the IRB review process by the KGCP regulation. Since the KGCP is the only regulation to obligate IRB review, the regulation does not apply to all the other researches that do not seek KFDA approval. The survey showed that only 30% of the whole IRBs reviewed the protocol of the academic researches, which demon- strated that there is room for improving the current IRB review system for better protecting research participants in academic studies (Table 4). The remaining mechanism of human subject protection is the publication requirement of the ethics review on topics involving the participation of human subjects, but these are not effective due to a lack of compliance by researchers in Korea. Recently, positive signs appeared for increasing the coverage of IRB review system. The Stem Cell Research Project supported by the Korean Ministry of Science and Technology, starting October 2002, set internal guidelines that all research protocols funded by the Project should be reviewed by the ethics committee and by the IRB of each institution where the research is conducted. In terms of research that is both socially and ethically controversial, such as stem cell research, the establishment of an ethics review system will increase the transparency of research practice, and thereby better protect the public. Second, few research institutions other than hospitals have IRBs. More than half of the IRBs in Korea were established in the hospitals accredited in 1993 for clinical trials immediately before the implementation of KGCP in 1995. Many other institutions conducting human participant research or equivalent research in universities or biomedical research centers do not have IRBs yet. However, the researchers in biomedical research field in Korea have recognized the importance of ethical consideration for human related research. This will evoke the establishment of new IRBs at biomedical research institutions other than hospitals in the near future. Third, most IRBs suffer from lack of independence. The very heart of the ethics review is the independence of the IRBs from political, institutional, professional, and market influences in terms of their composition, procedures, and decisionmaking (7). In 91% of IRBs, the president of the hospital appoints members, and in more than two thirds of IRBs, the chairperson is appointed by the president (n=29, 46%) or the president automatically takes the chair (n=19, 30.2%). As Wood et al. (8) pointed out, the IRB has inherent institutional conflicts of interest because most IRB members work for the very institution conducting the research they review. This makes the IRB and individual members inclined to be influenced by the senior officials and their peer relationships with researchers within the institution. Fourth, IRBs are not properly balanced in their membership constitution. To conduct competent and independent reviews, it is crucial to have IRBs that are soundly composed. An IRB should include relevant scientific and ethical expertise and laypersons, with balanced professional, age and gender distributions (8). Our survey shows that with the exception of male physicians (50% of all members), IRB compositions do not reflect diversity in terms of the representations of professions or genders. The most serious problem is that only 43% have layperson members and about 40% do not have members who are not affiliated to the institute. This may lead to a weakening of IRB independence and to depriving the community of research awareness. Finally, yet importantly, IRB members and investigators do not get adequate education. Although the KGCP requires the regular education of IRB members and clinical investigators, our survey shows that less than 20% of institutions provide regular education to IRB members and only about 10% to investigators (Table 2). Some institutes provide annual workshop programs and courses on human research for researchers and IRB members, but there is no specified standard curriculum available for training ethical research practice, even in these institutions.

Problems in the review process

The survey results showed significant variability in terms of the quality and expertise of IRB reviews. A small number of IRBs have members with more diverse expertise for reviewing the various protocols, while the remainders do not. Review practice itself varies among the IRBs. While members read and reviewed documents before a meeting in 41% of IRBs, the document review is replaced by an investigator’s presentation in 14% of IRBs. As the respondents replied (Fig. 5), the lack of experience and expertise of the IRB review and operation are the most serious problems for many IRBs. Although the KGCP, revised in 2001, addresses continuing review, only half of the responding IRBs conduct continuing review. Since the questionnaire was not detailed enough to gather information as to how continuing review is conducted, we cannot judge how effectively and efficiently the continuing reviews are conducted as an assuring system to protect research participants. As the concept of continuing review was recently introduced to the KGCP in 2001, several IRBs have been gearing toward installing continuing review that uses electronic database systems. While most IRBs in Korea have an expedite review system to deal with reports on serious adverse events or minor protocol changes, about 10% do not have expedite review system, which may cause time-consuming, inefficient review process. The categories of expedite review are also variable. Even 40% IRBs use an expedite review system to review the amendment of protocols that may increase risks to participants, or exert serious impact to clinical trial, which itself should be reviewed at a regular IRB meeting. These results (Table 8) suggest that specific guidelines on expedite review should be prepared.

Policy problems related to the protection of research participants

One of the most important functions of the IRB is to ensure the protection of research participants, which includes reviewing and overseeing the safety measures taken to protect participants. Although international guidelines suggest reviewing all kinds of compensation, such as treatment, prorated payments, indemnity, gifts, and other benefits, the survey shows that most IRBs review only for monetary payment. About 10% of IRBs do not review any kind of compensation for participants. In case of clinical trials, about 50% of institutions make sponsors indemnify against juries or harm to participants. Only 22 IRBs (35%) make it obligatory for sponsors to have insurance indemnity for participant injury. Except 3 institutions that have no policy, most have some kind of policy that covers participants in clinical trials, mainly because sponsors provide the necessary funding. More serious than the case for clinical trials is the lack of policies in place for the protection of participants in academic research. Currently, this has become more obvious and more serious, as more academic researchers face the problem. The financial resources required to indemnity participants in academic research are not as available as they are in clinical trials. Usually, research funds come from the government or public research institutes, which have tight research budgets. Twothirds of institutions do not have any policy for the indemnification of research participants. In 21% institutions, principal investigators have to pay for indemnity; in fact, in 10% of institutions they have the responsibility to do so. To date no insurance system for academic research has been developed, and problems in this quarter will undoubtedly grow.

SUGGESTIONS AND CONCLUSIONS

In spite of the problems with the IRB system, these data are neither complete nor without error. Since this study relied solely on the reply of an expert member of each IRB to the questionnaire, further evaluation by an independent third party is necessary for the systematic collection of thorough and valid data. To improve the current situation of human research participant protection in Korea, We make the following suggestions. In order to provide the public with assurance, it is necessary to make regulations to obligate all human research categories to be reviewed by research ethics review committees. In spite of the explosive growth of human resource research in Korea, there exists only a regulation concerning research requiring KFDA approval (KGCP). At least, governmentfunded research, which is virtually citizen and taxpayer financed, should be reviewed by IRBs to ensure the protection of research participants. To improve the quality of the IRB review, it is important to build networks at regional, national, and international levels, to exchange information, utilize resources efficiently, and standardize the quality of review. KAIRB is expected to play important supportive role for local IRBs, by regularly providing necessary information and by educating IRB members. KARIB has recently developed general guidelines for the constitution and operation of IRBs, which encompass hitherto uncovered areas of human participant research. There is also a need to organize regional IRB meetings in Korea, and to create joint IRBs to adequately manage the requirements of multi-centered research. At the international level, the WHO is organizing regional forums of ethics committees charged to review biomedical research (9). To improve the quality of the IRB review system, continuing education and training is essential. As several important reports, as well as our survey respondents pointed out, one of the serious problems of facing IRBs in developing countries, including Korea, is lack of expertise in the IRBs. Primarily individual institutions ought to take the responsibility for training IRB members and the investigators, in ethics of research involving human participants. Realistically speaking, when lack of resources at an individual institute becomes a real problem, the KAIRB, relevant academic societies, the KFDA, or some other responsible governmental organizations should be able to develop and provide curricula and course programs, which include basic, advanced, and specific courses for all IRB members and investigators. Eventually, we need to build a quality assurance system (accreditation) for research institutions. As more researchers now recognize the importance of the IRB review, it is timely to consider an accreditation system requiring the continuing selfassessment, self-improvement, and auditing of IRBs. With the globalization of biomedical research, the number of international agencies overseeing IRBs will increase through auditing or inspection (9-12). As a part of the peer review or for educational purposes, the auditing or surveying of IRBs should be introduced in Korea. After all, these activities will demonstrate the continuous efforts made by our Korean IRBs to build capacity for conducting biomedical research of the highest attainable quality in terms of both science and ethics.

With the Digital Revolution, the time-honoured model of scientific discovery being contingent on a singular intellect working independently of others has expired. In the modern age of global travel and the interactive capabilities afforded by the internet, there is an expectation that good researchers are internationally mobile, both physically and virtually (Schiermeier, 2011). Researcher mobility is not a goal in itself, but rather a means of fostering collaborative networks at the many levels (e.g., institutional, interdisciplinary, international, etc.) that may drive successful scientific discovery. The increasing dominance of collaborative teams both within and between institutions has been documented to enhance efficiency and productivity as well as produce better science (Wuchty, Jones & Uzzi, 2007). This is also reflected in the growth of international teams and their association with increased citation counts, a marker of research impact (Wuchty, Jones & Uzzi, 2007; Jones, Wuchty & Uzzi, 2008) Entangled within this collaborative research milieu, the institutional affiliations held by a researcher may also be viewed as a marker of capacity to facilitate knowledge exchange (ESF, 2013). However, to date there has been little research from the burgeoning scientometric and bibliometric fields exploring the role of multiple institutional affiliations on scientific output (Hottenrott & Lawson, 2017). To improve our understanding of this phenomenon, we conducted a large-scale analysis of scientific publications from four multi-disciplinary science journals (Science, Nature, Proceedings of the National Academy of Sciences (PNAS), PLOS Biology (PLOS)).

MATERIALS & METHODS

We retrieved all `articles' listed for the above journals from Web of Science (WoS) for the years 20102014, inclusive (search performed on 14/06/17). Articles were exported from WoS as BibTeX files, with complete metadata, then imported into the R statistical environment (R Core Team, 2017) for further processing. The bibliometrix package (Aria & Cuccurullo, 2016) was used to create a bibliographic data frame with cases (rows) corresponding to manuscripts and variables (columns) to Field Tags (metadata) in the original BibTex file. In this way the bibliographic attributes for each article (i.e., title, author's names, author's affiliations, citation count, document type, keywords, etc.) are formatted appropriately for subsequent analysis. The most important Field Tag for the purposes of this study is the Author Address (C1) tag which provides institutional address information for each author and where an author has multiple affiliations, lists these addresses separately. We split each manuscript record by author name and affiliation address, with the sum of author name occurrences indicating the number of distinct affiliations for that author. As comparisons of raw citation counts are biased by virtue of time since publication (i.e., earlier publications have had longer to accumulate citations), normalized citation counts were computed by dividing the raw value by the number of days since June 30th of the year of publication through to the search date (14/06/17), and then multiplying by 365 (Uddin & Khan, 2016). This enables unbiased comparisons of citation counts irrespective of the year of publication.

RESULTS AND DISCUSSION

Of the 27,651 articles retrieved, 39 did not have affiliation data recorded and were excluded. The total number of articles available for analysis was 27,612, with Science (nD3,910), Nature (nD4,120), PNAS (nD18,651), and PLOS (nD931). The maximum number of citations for a single paper (published in 2012) was 4,143 (mean and median: 79.6 and 43.0, respectively). The maximum number of normalized citations was 828, for the same paper (mean and median: 15.7 and 8.8, respectively). The maximum number of authors for a single paper was 2,908 (mean and median: 9.0 and 6.0, respectively), and the maximum number of author affiliations was 271 (mean and median: 4.7 and 4.0, respectively). Author affiliations were recorded as presented by WoS.

Table 1 shows the distribution of article and author appearances stratified by the number of author affiliations for the most- and least-cited articles split at the median normalized citation value (Highest citationsDcitations>8.8 [nD13,795], Lowest citationsDcitations 8.8 [nD13,817]). While the vast majority of author appearances were associated with only one institutional affiliation (74.1%), 25.9% of author appearances were linked with two (20.0%) or more affiliation addresses. The maximum number of institutional affiliations held by an author was 12. As these are non-independent observations, classical tests of contingency tables are not appropriate; however, one can easily appreciate the increased frequency of author appearances in the more-cited publications. Indeed, the correlation between the normalized number of citations a paper received and the number of authors on that paper was statistically significant (D0:17, p0:001). Similarly, the correlation coefficient for the normalized citations a paper received and the number of instiutional affiliations on that paper was 0.25, p 0:001. The correlation between the number of authors and number of affiliations listed for each paper was greater, indicating closer correspondence between the variables (0.67, p0:001). To facilitate a simple yet fruitful investigation of the relationship between the number of normalized citations a paper received and its association with authorship and affiliation frequency, we categorised the latter two variables. The number of authors attached to each paper was split into quartiles to create an `Author Number' variable, with the following categories: 1D13 authors/article, 2D45 authors/article, 3D69 authors/article, and 4D102,908 authors/article. Due to the low cell counts (Table 1) and to improve estimation in subsequent modelling, the maximum number of author affiliations held on a single paper was limited to six. This resulted in the exclusion of a further 47 papers, with 27,565 articles available for analysis. `Maximum Affiliation' represents the maximum number of institutional affiliations held by a single author on an article. For example, if WoS listed an article with three authors each having two affiliations, and two authors each having three affiliations, in this case maximum affiliation would equal three. Table 2 shows the frequency distribution of articles by author number and maximum affiliation. Figure 1 shows boxplots of citation counts for each category of author number and maximum affiliation. There is a general trend of normalized citation count increasing across both factors. We explored this relationship further in a linear regression model with normalized citation count as the outcome, and author number and maximum affiliation as predictor variables (Table S1). Although these are technically count data, the mean citation value is high and the distribution of the count model approximates the normal. Consequently, we have considered citations a continous variable and utilised a linear model. We initially fit a model with an interaction term (author number maximum affiliation) and evaluated its signficance with a Wald test. The resulting p-value was highly significant (<0.001) suggesting the 15 coefficients for the interaction terms are not simultaneously equal to zero, and an interaction effect exists between the two variables (i.e., the relationship between maximum affiliation and citations received, varies depending on the value of author number). The model was checked for multicollinearity using the generalized variance inflation factor (GVIF). The raw output from the regression model is supplied in the Table S1. As interaction terms make coefficient interpretation difficult, results for the effect of each level of predictor are presented in a stratified manner, while holding the other predictor constant (Table 3). In addition, we adjusted for year of publication and journal in the analysis. It is of interest to note the effect of journal on normalized citation counts. Using PNAS as the reference category journal (chosen as the most populous), both Science and Nature receive on average higher normalized citation counts per paper (p < 0:001) in comparison. Citations received were not significantly different between PNAS and PLOS. Table 3 shows the effect for each combination of maximum affiliation and author number on normalized citation count. To further facilitate interpretation, we have limited maximum affiliation data to four addresses. The effect size (average change in normalised citation count) was computed using a series of linear contrasts that enables the comparison of differences among coefficients beyond the standard regression output. There are two main findings from these data: first, the effect on citation count of an author holding more institutional affiliations increases as the number of authors on a paper grows; and second, increasing the number of authors on a paper tends to result in more citations received irrespective of the number of affiliations held. When there are between one to five authors/article, increasing the number of affiliations an author holds (relative to one) does not affect the average change in citation count. However, when there are between six to nine authors/article, authors with two institutional affiliations (relative to one) will, on average, increase the citations a paper receives by 1.6 (p D 0:006). This effect is even more pronounced when there are more than nine authors listed; here, citations increase on average by 2.3 (pD0:002) for two affiliations, 5.8 (p<0:001) for three affiliations and 9.4 (p<0:001) for four affiliations, relative to the reference group. If we now interpret these effects while holding the number of affiliations constant, for researchers with only one affiliation, increasing the number of authors on a paper results in a mean increase in the citations received across all levels of author number (e.g., 6.5 for author number D 4, relative to 1, p<0:001). However, this effect remains significant for only greater author numbers (i.e., four vs. one) as the maximum number of affiliations held, increases. We would like to remind the reader that these data are cross-sectional in nature, and our discussion of `effects' in the context of regression analysis does not imply causation in the relationships explored.

CONCLUSIONS

These data align with previous observations in highlighting the increasing leverage of teamwork in scientific research (Wuchty, Jones & Uzzi, 2007; Jones, Wuchty & Uzzi, 2008). They also serve to provide some insight into the relatively novel notion that multiple author affiliations may play a positive role in the production of high-impact science (Hottenrott & Lawson, 2017). However, longitudinal analyses of citation count data would be necessary to explore the basis for a causal relationship. To that end, further research is needed to address some of the questions arising from the main finding of this study. What causes multi-institutional, larger authored papers to have greater citation impact? Is increased institutional representation seminal in the generation of high-quality science and therefore more highly cited works? Or are we observing an artefact of highly-funded and highly-competitive research that by its nature will generate more citations, irrespective of the number of authors or their affiliations. Clearly more data are needed to comprehensively address these points. Until then, the holding of multiple affiliations by authors should be viewed by institutional boards as a virtue and not a vice, as it appears that greater researcher mobility may be advantageous to all.

For the past few decades, biomedical research has undergone a process of globalization, with a remarkable increase in the amount of research conducted in low and middle-income countries (LMICs)(1). To further this progress, there has been a push to build research capacity, especially across countries in Africa. In recent years, several international organizations such as the African Institute of Biomedical Science and Technology, The Wellcome Trust, and the World Health Organization began initiatives aimed at increasing research within African institutions and improving training for African scientists(2-4). This growth also generated a parallel expansion of efforts to address the ethical aspects of international research. That is, as research capacity increases, questions such as those relating to priority setting, informed consent, and the ability of individuals and institutions to monitor research also arise(1). To help address this growing need, several programs have emerged to help researchers and research institutions in LMICs increase their capacity in research ethics. These programs include those from the Wellcome Trust, Fogarty International Center of the U.S. National Institutes of Health, and the European Developing Countries Clinical Trial Partnership(5-7). The goals of these programs are often to support the development of research ethics capacity in order to enhance research oversight and research ethics systems, scholarship on locally pertinent topics in research ethics, and locally developed training opportunities(8). As capacity in research ethics grows, however, it is also prudent to examine where and how programs are targeting their work to ensure that they are tailored to local needs, are properly implemented and significant progress is being made. This allows LMIC researchers and institutions to identify specific targets for improvement. Evaluation is a key part of capacity development in any field; the growth of such evaluations has allowed for more targeted and successful research interventions( 9). However, very few publications have focused specifically on evaluating research ethics and bioethics programs(10, 11). Importantly, many efforts to increase research ethics capacity have focused on training individuals and research ethics committees; however, there is also a need to focus on building institutional research ethics capacity to further enable and sustain a culture of ethics(12). As will be discussed further, the “systems” approach that we used in this case study is one example of this type of framework. It is important to note that research ethics systems can be influenced by outside factors such as wider community values, regional or national regulations, and international norms(13).

The Training Program and Institutions

The Johns Hopkins-Fogarty African Bioethics Training Program In 2012, The Johns Hopkins-Fogarty African Bioethics Training Program (FABTP) began a one-year institutional partnership with the University of Zambia (UNZA), and specifically with the School of Medicine through the Department of Public Health. The history of FABTP and its partnership model have been described previously in several publications(10, 14). The primary goal of the partnership was to help further research ethics capacity within the university. In order to proceed with meaningful engagement and support future benchmarking, a systematic approach to assess baseline institutional research ethics capacity and needs was initiated. In this paper, we use the UNZA case study to demonstrate a rapid approach to assessing institutional research ethics capacity within the context of an LMIC. Since 2010, FABTP has completed two such evaluations at universities in Uganda and Botswana(10, 14). The UNZA case study employed many of the same methods, though it builds on the previously conducted assessments by attempting to produce a rapid version of an institutional assessment. Our aims are to present the baseline evaluation of research ethics capacity at UNZA, as well as demonstrating the utility of the framework for assessing institutional bioethics development capacity. We begin with background information on the university as a whole and the School of Medicine, then discuss the specific model used for the assessment, followed by a discussion of the application of the model in this case study. We aim to demonstrate how the model used in this case study can be applied in other contexts as well.

The University of Zambia

UNZA, founded in 1966, is one of six public universities in Zambia and matriculates just under 10,000 students, the great majority of whom are undergraduates(15 ,4). The university is comprised of ten different schools including the School of Medicine, which was founded independently in 1965 and joined UNZA in 1970(15). In addition to training health professionals, the school also conducts biomedical and public health research. The School of Medicine began offering graduate Master in Medicine degrees for clinical disciplines in 1985 and a Master of Public Health degree in 1995(15). The commencement of these graduate programs increased research volume in the university in general; and this resulted in calls for capacity strengthening in both health research and research ethics. As a research university, UNZA incorporates ethics into both its teaching and research practice. As will be discussed further, at the time of the baseline evaluation, these responsibilities fell into three main university domains: the research ethics curriculum, the Directorate of Research and Graduate Studies (commonly referred to as, “The Directorate”), and the Biomedical Research Ethics Committee (hereafter referred to as, “the ethics committee”). Each of these addressed a different facet of research ethics and was the focus of this evaluation. The research ethics courses at the university were coordinated through the School of Humanities and Social Sciences and were available for both undergraduate and graduate students. The Directorate was responsible for coordinating research, developing research policies, and supporting student research. The ethics committee, established in 1982, was tasked with reviewing and approving all biomedical research protocols including any protocols involving human or vertebrate animal subjects(16).

At the time of data collection, the university also operated two other research ethics committees in addition to the above mentioned committee; one focused on natural and applied sciences and another for social sciences(16). These two committees were not assessed in this evaluation.

Methods - Approach to Institutional Capacity Assessment

We took a “systems” approach to our evaluation of research capacity. This meant focusing on the institution as a whole – evaluating the research ethics programs, as well as the organization and infrastructure of the institution. As we have detailed in previous publications, institutional research ethics systems should include the following: 1) a justifiable research agenda, 2) protection of research participants, 3) training for institutional members, 4) the creation of institutional priorities and structures that promote ethical conduct, and 5) strengthening communications with regional, national and international stakeholders, especially institutional leadership(13). We conducted a rapid assessment utilizing several methods to elucidate the breadth and scope of the university’s research ethics capacity. First, an institutional survey was administered to the local partnership leaders (faculty within the School of Medicine) who then gathered information from other university faculty as needed. The survey contained 168 questions on a range of topics including educational opportunities, training for committee members and finances. Individual questionnaires were also administered to members of the ethics committee with questions regarding their specific training and role within the committee (Table 1). The evaluation also included a site visit by the FABTP partnership team to the School of Medicine, which took place in January 2012, as well as a document review to better understand bioethics programs, institutional policies and future plans. We used the Octagon Model for our evaluation, which was developed by the Swedish International Development Cooperation Agency(17). As designed, this model is intended to evaluate nongovernmental organizations in order determine their strengths and weaknesses(17). The model details eight domains (hence “octagon”) that are used to help provide an overall analysis of the organization: basic values and identity, structure and organization of activities, implementation of activities, relevance, right skills in relation to activities, systems for financing and administration, target groups, and working environment (Table 2). Each of these eight domains is rated on a scale of 1 (the lowest) to 7 (the highest) (Table 2). We have previously detailed how we have adapted the Octagon Model for the institutional research ethics context, using modified criteria(10, 14).

Table 2: Ranking scale and description of the Octagon Model domains as applied to institutional research ethics evaluation

Our evaluation and octagon score were based specifically on assessing the School of Medicine, The Directorate, and the ethics committee. Two FABTP faculty members ranked each of the eight categories after reviewing all of the data. UNZA faculty members also prepared an octagon score based on their own perception of the institution and the two scores were compared.

Results - Application of Framework

Basic Values and Identity:

The School of Medicine Strategic Plan for Operations was finalized in 2012 to help establish the school’s goals and vision. While the document did discuss the school’s research priorities, the plan only briefly mentioned research ethics, in the context of the school’s objective to further develop ethics capacity and in stating the existence of the ethics committee(18). There were no further details regarding how the school planed to meet its ethics objectives. Further incorporating research ethics into the school’s strategic plan could help establish ethics as a core element of the school’s research program. School of Medicine leaders involved in bioethics program planning and implementation stated during conversations with FABTP staff that their vision was to expand the scope of bioethics within the school in the near future, and improving research ethics training for faculty was a top priority. They spoke of plans to create a dedicated bioethics center that will reach across the school, and become a formal home for bioethics teaching, research and service within the entire university. These goals demonstrated that integrating bioethics within university teaching and research practice were key priorities for UNZA. Structure and organization of research ethics activities: We learned from speaking with university faculty and reviewing school documents that research ethics within the university fell under the responsibility of several different individuals and departments. There was an Assistant Dean of Research who was responsible for overseeing research, as well as the Director of Research and Graduate Studies who oversaw all research ethics activities Additionally, The Directorate, an office of the graduate school, was responsible for coordinating research, developing research practice and supporting student research work. The office reported to the Vice-Chancellor of the university and was comprised of sixteen members, two academic staff and fourteen support staff. However, at the time of data collection, The Directorate did not provide a written document outlining its mission or a formal framework for the functions of the office. There was an organizational chart depicting the chain of management for research; however, there was no depiction of how the office related to other units and departments within the university. Such a chart may help clarify the role and responsibilities of The Directorate. The structure and management of the ethics committee was detailed within the committee’s standard operating procedures. There were clear instructions for membership requirements and leadership positions. The committee was made up of a diverse group of members, with representatives from the School of Medicine, University Teaching Hospital, and the Ministry of Health, amongst others. These members represented a range of training backgrounds, and two of the members were external and not employed by the university. However, though the standard operating procedures outlined the roles of the committee members, it was not clear where the committee sat administratively within the rest of university or the School of Medicine. Responses within the evaluation questionnaire indicated that formal academic courses in research ethics were offered through the School of Humanities and Social Sciences only. These classes were offered on campus for registered students. As of 2012, research ethics classes were not offered within the School of Medicine or School of Natural Sciences. Faculty members were responsible for securing their own ethics training at workshops or courses that took place off campus, including multi-day trainings at universities in other countries. There was some funding available for faculty to attend such trainings. Implementation of research ethics activities: As mentioned above, responses from the questionnaire indicated that UNZA offered several research ethics educational opportunities through the School of Humanities and Social Sciences. These included tutorials, academic lectures, and a sub-concentration in bioethics for Masters’ students. Students conducting research must have completed coursework in research ethics as well, though there was no general requirement for all students to complete these courses. However, it should be noted that, according to the questionnaire, several critical ethics-related topics were absent from the courses offered to students. These included: informed consent, human rights, and research with vulnerable populations. Additionally, there was no requirement for UNZA researchers to complete any form of training in research ethics. UNZA faculty stated that while the role of overseeing research within the university was officially the responsibility of The Directorate, in practice this was mainly handled at a departmental level. It was the departmental chairs and academic faculty who guided students and coordinated research within each department. Faculty suggested that the formation of an independent bioethics department or center may help raise awareness for the importance of ethics more broadly and serve as a resource for students, department heads and others. The ethics committee, in order to accomplish its goal of providing ethics review of proposed human and vertebrate animal research, met once a month to review research protocols. As indicated in the answers to the questionnaire, in total, the committee evaluated 127 protocols in 2011, approximately 100 of which were local studies (i.e., no international collaborators or funding involved). An expedited review process was used for approximately 30% of protocols. In these cases, the chairperson was principally responsible for reviewing the proposals though it was unclear what criteria were used to determine whether a protocol qualified for expedited review. Additionally, the ethics committee was registered with the U.S. Office of Human Research Protection and had a Federal-wide Assurance. Though the committee had documentation for its administrative structure and protocols, there was a need for improved implementation of these policies and improved efficiency. UNZA staff stated during interviews that the work of the committee was hampered by perceived inadequacies in communication between the committee members and researchers, and insufficient administrative staffing within the office. Since the committee was not housed within a specific institutional department or unit, it was unclear which entity was responsible for facilitating staff hiring. It was also reported that the ethics committee did not have a dedicated meeting room to conduct protocol reviews. These elements reduced efficiency within the work environment and made it difficult, at times, for the committee to execute all of its functions. Right Skills in Relation to Activities: As stated, the university’s leadership expressed that training faculty in bioethics and research ethics was an institutional priority. In conversations with FABTP members, UNZA faculty stated that they felt that, as of 2012, many university researchers lacked adequate knowledge of scientific study design and research theory. Strengthening training programs for researchers will not only improve the quality of the research within the university, but will also help expedite the work of the ethics committee. The committee was hampered by inadequate research proposals, largely due to the fact that researchers were not adequately trained in conducting studies and preparing such proposals. The ethics committee provided training for its members in the form of a three-day workshop; continuing committee members also completed this workshop in order to ensure that their training was up to date. Trainings were held every 1-2 years as committee members’ tenure came to and end and new members joined the committee. Additionally, several members of the committee, including the chairperson, had completed postgraduate training in research ethics. However, the ethics committee did not appear to have written training policies; formalizing training protocols will help ensure the productivity and stability of the committee. As detailed in the questionnaire responses, training was required for Directorate key personnel; however, the office itself did not offer any training programs in research ethics. Staff members were required to have completed a two day in-person training on the administration of ethics committees. However, it is unclear what topics were covered in this training, and if the focus was solely on ethics committees or included other topics. Relevance: As of 2012, UNZA had developed a vision for the future, though it had not yet begun to implement programs necessary to achieve its new bioethics goals. As stated, the university’s main goals were to strengthen its bioethics teaching and training programs, as well as create an independent department or center for bioethics. The university did have several elements already in place to facilitate pursuit of these goals, such as The Directorate and various deans keenly interested in expanding bioethics teaching and training. This existing framework can be utilized to ensure that UNZA carries out its plans for the future; however, as of 2012, this work had not yet started. One roadblock to fulfilling the school’s needs was the high volume of external research proposals submitted to the ethics committee. University faculty stated, and survey data confirmed, that the majority of the proposals reviewed by the committee (approx. 54%) were projects that were not affiliated with the university. Though the committee is meant to serve the greater Zambian research community, this large volume of work limited the committee’s ability to serve researchers based at the university. However, in 2011 the committee received a grant from the European and Developing Countries Clinical Trials Partnership, which, at the time of this evaluation, it was hoped would further research ethics development. Systems for financing and administration: University leaders highlighted that the lack of funding for ethics programs as a major challenge for building research ethics capacity. The university did not have any budget set aside for research ethics related activities. Thus, The Directorate operated without any dedicated funding for its ethics programs. The ethics committee did have a small operating budget and charged fees for reviews in order to raise more. The committee had a full time secretary seconded to it by the Tropical Gastroenterology and Nutrition Research Group, which was funded by external donors. Target Groups: UNZA ethics leadership recognized students and academic faculty as their main target audiences for bioethics programs. However, since individual departments were largely responsible for overseeing research and managing bioethics training, they too were stakeholders. Additionally, the work of the university also affected the larger research community both within Zambia and beyond. These groups must be taken into account when considering the bioethics programs within UNZA. As of 2012, students were not represented within the planning or implementation of ethics programs though they represented the primary target group for many of the university programs. Incorporating student input will help tailor programs to student needs and will increase program impact. Student input may also be helpful for crafting new ethics courses and improving existing courses. Strengthening the capacity of The Directorate will improve its ability to coordinate with stakeholders such as departments and faculty. Codifying protocols may also help standardize and clarify the role of The Directorate to the research community within UNZA. During meetings with FABTP staff, university members highlighted the need to improve communication between the ethics committee and faculty researchers. The committee did have clear instructions regarding what types of proposals must be submitted for review so researchers were aware of the committee standards. However, the distinction between the three research ethics committees within UNZA was unclear. Faculty members wrote in an internal document overviewing research at the university, there was “no clear delineation as to which of the three committees specific proposals must be sent to”(16). This lack of clarity not only made it more difficult for researchers to submit proposals but it also confused the role of each committee within the university. UNZA partnered with several other universities and organizations to expand the scope its ethics research and training programs. For example, the university maintained a relationship with Tan- ZamBo, a research collaboration between three universities within Tanzania, Zambia and Botswana. UNZA also worked with several other sub- Saharan universities as well as universities in the United Kingdom and the United States through the Southern African Consortium for Research Excellence, an initiative aimed at increasing research and funded by the Wellcome Trust(19). At the time of this evaluation, UNZA did not have a formal working relationship with the World Health Organization (WHO), or specifically with their regional office for Africa, with regards to bioethics programs. However, the School of Medicine did have several collaborations with the WHO on other health related projects. Working Environment: UNZA operates within the larger research environment of Zambia. As of 2012, there were national guidelines regarding research with human subjects, though there were no national laws regarding research ethics. There has been progress in the last several years, including the formation of the National Health Research Ethics Committee (NHREC), a government operated ethics committee( 20). However, within Zambia, UNZA acted as one of the leaders for research ethics and the ethics committee at the university was one of only a handful within the country and therefore was responsible for reviewing materials from all over Zambia. In addition to the national committee, there were ethics committees housed at the Tropical Diseases Research Centre, Macha Mission Hospital as well as the Excellence in Research Ethics and Science Converge Committee, a privately operated research ethics committee(21, 22). In 2013, the parliament passed the National Health Research Act, which included the formation of the National Health Research Ethics Board(23). The board was tasked with generally overseeing research and ethics programs within the country as well as the abovementioned national committee(23). This Act, and the board specifically, helped establish a national framework for research ethics. Though new, it will hopefully provide support for UNZA and create a broader cooperative community for research ethics development within the country. Overall Assessment: To supplement the above, primarily qualitative, descriptions, we also used the Octagon framework to generate overall octagon scores tabulated after completion of the baseline evaluation. Figure 1a is the score prepared by the faculty and staff at UNZA; Figure 1b is the score prepared by members of FABTP. Two members of the FABTP team independently reviewed data and created scores in all eight categories; then differences between scores were discussed and reconciled to generate a final FABTP score. The UNZA team also separately convened to consider data and generate a consensus score. The FABTP octagon score recognizes strengths in the working environment and right skills domains. Systems for finance and structure score highly as well. The area of most need is relevance, followed by target groups and identity. The UNZA self-assessment demonstrates that the university is confident in its work in several areas: basic values and identity, systems for financing, right skills and working environment. These scores demonstrate that UNZA self-identifies areas for improvement in all categories, though no single domain is weakest and each ranks at least at a “good” level. The discrepancies between the internal and external assessments are important to note as these areas warrant further discussion and investigation.

Discussion

In this paper, we describe a rapid assessment to evaluate the strengths and weaknesses of an institutional research ethics system using a novel approach. This case study builds on previous work by demonstrating a faster, less resource-intensive approach to program evaluation. Our assessment focused on The University of Zambia, and specifically on the School of Medicine. We evaluated the programs individually as well as within the greater context of the Zambian research environment in 2012. We believe that this approach provides a useful initial overall evaluation of UNZA as a research institution that is pursuing enhanced capacities in research ethics. When entering this partnership with FABTP, UNZA identified four main goals for the year. These were: 1) to evaluate the research ethics capacity of the university; 2) to create a bioethics unit that works across all departments within the School of Medicine; 3) to strengthen ethics teaching, research and service capacity by training public health faculty in research ethics; and 4) to create ongoing research ethics programs for faculty and students focused on training, scholarship and mentorship. These goals demonstrated that building both individual and institutional bioethics capacity was a priority for UNZA and these tangible goals focused on the needs of the university. Our assessment highlighted several strengths of the UNZA bioethics program, as well as particular areas where additional focused effort will likely help the university reach its goals. The university had several faculty members who were well trained in research ethics and committed to improving the research ethics capacity. The university also acted as a leader within the broader Zambian research community and expanding research capacity within the school will strengthen this role. In order to achieve the goals that it has set for itself, and meet the needs of various internal and external stakeholders, UNZA needs to ensure that the programs it implements are relevant. Clarifying the roles of the various offices that are responsible for research ethics will help standardize and strengthen each of these departments. A responsibility of one such office could be to con duct additional and ongoing stakeholder engagement and needs assessments to ensure connectivity between activities and audiences. Including research ethics within the university’s strategic plan for research may also help solidify the role of ethics within UNZAs infrastructure. A comparison of the external and internal assessments was most useful for our joint discussions between FABTP and UNZA (Fig. 1). The general similarity of the octagons overall demonstrated some consistency of outcomes, while the differences reflected either variations in perceptions or understanding of the approach – though all involved in scoring the categories used the same rubric and guidelines in order to reduce scoring errors. It is also notable that external and internal assessments for research ethics capacity were most divergent around what were more generally the weaker aspects of the institutional research ethics system. That is to say, both UNZA and FABTP identified similar areas for improvement, though scores attributed by FABTP were somewhat lower in those areas.

We believe that the Octagon Model is a useful tool for evaluating baseline program capacity, even though it has limitations. As employed in this assessment, the approach was conducive to understanding general (macro-level) information on various topics; however, more detailed information was often left out. For example, though the questionnaires gathered information on courses offered in research ethics, there was no information as to the quality of these courses or how many students enrolled in them. Additionally, in this case, the assessment did not reflect the opinions of all stakeholders within UNZA. Most notably, we were unable to speak with university students. In this rapid assessment, we did not conduct specific focus groups or in-depth interviews with students and faculty as we have in previous case studies(10, 14). These tools allow us to gain more information and a wider range of opinions regarding ethics capacity(10, 14). Despite these limitations, we believe the approach used provides a reasonable means to initiate empirically-informed institutional capacity development in bioethics, and perhaps more broadly. A limited number of tools are available for evaluating research ethics capacity, especially within LMICs; with even fewer incorporating institutional and other contextual elements into the assessment(24). The Octagon Model provides a multi-level framework and allows institutions to use data specifically to target efforts in domains that are typically administratively and organizationally meaningful. This case study in particular demonstrates the utility of an assessment that can be performed with fewer resources and in less time as compared with other evaluations. We believe that this type of rapid systematic approach will be helpful for further evaluations and should be applied to other institution in order to refine the approach and make it more user friendly in the future.

The remarkable progress in computational and communication technologies in recent years has made it possible for scientific and technical data and information to be produced and reused at an ever-accelerating rate. This has started to have a transformational impact on research activities in science (Emmott 2006). These novel fields of scientific research increasingly depend on the collection, transmission, and utilization of a vast amount of information and knowledge. Electronic infrastructure established for this function has come to be referred to as ‘Cyberinfrastructure’ in the US, ‘e-Science’ in Europe, and ‘e-Infrastructure’ in Japan (David et al. 2003; National Science Foundation 2003). As the data intensity of research activities in scientific fields has increased significantly, the sharing of information and knowledge is expected to provide a critical foundation for accelerating scientific and technological development. These developments in scientific research has enabled the rapid creation of information and knowledge and easy access to their sources, which are essential components of innovation (Foray 2004). Especially in fields where scientific progress is rapidly developing and the sources of information and knowledge are widely distributed, no single individual or organization has all of the necessary skills to stay on top of all areas of progress (Powell and Grodal 2005). Previous research confirms the important role of external sources of scientific information in bringing forth significant breakthroughs (Freeman 1991). Dense ties between partners in collaboration networks contribute to fostering information diffusion and knowledge exchange, enhancing the scientific performance and collaborating opportunities of the partners (Uzzi 1997; Stuart 1998; Ahuja 2000; Yarime 2009a, b; Baba et al. 2009). Recently, we have observed new scientific fields which transcend traditional boundaries of academic disciplines, that is, inter-disciplinary or trans-disciplinary science (Haberli and Klein 2001). Among the emerging fields of inter-disciplinary science are bio-informatics (Matsuda et al. 2006), material informatics (Chikyo 2006), complexity science, and sustainability science. Sustainability science, in particular, aimed at understanding the fundamental character of interactions between natural, human, and social systems, covers a wide range of academic disciplines (Kates et al. 2001; Clark and Dickson 2003; Komiyama and Takeuchi 2006). Since the challenge of sustainability is the reconciliation of society’s development goals with the planet’s environmental limits over the long term (Clark and Dickson 2003), it is of critical importance to make appropriate use of knowledge and information on diverse aspects, ranging from the natural environment and artifacts to economy and culture. Sustainability science, thus, needs to be based on a firm understanding of the fundamental characters of complex interactions and interdependencies between natural, human, and social systems at the global scale. This will require an integration of various academic disciplines, from natural sciences and engineering to social sciences and humanities. A recent study shows empirically that the academic landscape of sustainability science actually consists of clusters of different disciplines (Kajikawa et al. 2007). It will be crucial that scientific information and knowledge are created and communicated effectively as well as efficiently, transcending disciplinary and geographical boundaries in the field of sustainability science. When tackling cross-cutting problems, as particularly expected in sustainability science, assembling appropriate expertise could be one of the prime reasons for scientific collaboration (Shrum et al. 2007). In reality, however, scientific knowledge and information are not necessarily shared or integrated effectively beyond established organizational or institutional boundaries (Maurer 2006). This has particularly serious implications in the case of sustainability science, which deals with diverse types of discipline and expertise. There are technical, economic, policy, and legal/institutional barriers and obstacles discouraging collaboration in scientific activities. This paper is an attempt to examine quantitatively the patterns of collaboration in the emerging field of sustainability science. The patterns of scientific collaboration are analyzed by utilizing bibliometric data on scientific articles published in academic journals in terms of geographical boundaries and research subjects. Based on empirical findings on the current state of collaboration, opportunities as well as challenges in establishing research collaboration in sustainability science are discussed. Implications will also be drawn for organizational and institutional arrangements to be set up for future research.

Data and analytical methodology

It has been an issue of debate for what could be included in the field of sustainability science. As we have not yet reached a state of consensus on the definition, concepts, or methodologies which should be adopted in this emerging field of sustainability science, in this paper, we limit our investigation to examining scientific articles which contain the term ‘sustainability.’ This is basically the same approach taken in previous research which attempted to examine quantitatively the academic landscape of sustainability science (Kajikawa et al. 2007). An empirical analysis is conducted on the patterns of research collaboration by looking at the co-authorship of articles published in academic journals. An assumption is made here that knowledge creation and sharing in research is reflected in the co-authorship patterns of publications. More practically, a data set of research articles which include ‘sustainability’ or ‘sustainable’ in their titles, abstracts, and keywords is created. Data on these publications are collected from the Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI), compiled by the Institute for Scientific Information (ISI), two of the sources utilized frequently for citation data. This choice of databases limits the scope of our examination to scientific articles published in English. Although that would involve a certain degree of bias favoring articles written by researchers in Englishspeaking countries, currently, these two databases are the most reliable bibliometric sources, covering a very wide range of academic journals in diverse fields. As academic activities through international collaboration tend to be conducted in English, with the resulting scientific articles mainly published in English, we assume that the English bias would be negligible in examining the patterns of international collaboration. Articles are picked up by using ‘sustainab\*’ as the keyword of the query, where the \* means a wildcard so that the corpus built includes those papers that contain ‘sustainability’ or ‘sustainable.’ The retrieved data includes the bibliographic record of author affiliations. The frequencies in which the countries of the author affiliations appear, as well as that of the cooccurrence of them, are counted in the records. After gathering these statistics, the collaboration intensity of two countries, Inti,j, is calculated by the following equation:

where ni and nj are the number of papers of country i and j, respectively, and kij is the number of co-authored papers by country i and j. The number of links between the two countries is normalized by the number of papers written by each country in order to cancel out the convergence effect to specific countries which have a large number of papers. A high value of collaboration intensity suggests that the two countries tend to have a large degree of collaboration between them in their research publications. Also calculated is the collaboration-intensity similarity of two countries. It is calculated in the following way: Simi;j ¼ kij ki kij kj N2 ninj ; where ki and kj are the number of collaborating countries of country i and j, respectively, and N is the number of all papers. A pair with a high value of Simi,j means that the paired countries tend to focus on collaboration between them relative to other research partners. Then, a network is created based on the data of the collaboration-intensity similarity. We visualized the structures of collaboration networks by using the Fruchterman– Reingold (FR) method (Fruchterman and Reingold 1991), which is based on a spring layout algorithm where links play the role of spring-connecting nodes. The social network analysis tool Pajek was used for the visualization. As a result of such layout, a group of countries which write articles together tend to be located in closer positions. We regard Simi,j as similarity between the countries. The width of links between countries in the visualization is drawn to be proportional to Simi,j. The inter-disciplinary field of sustainability science requires, by its nature, diverse types of knowledge, reflecting the complexity of the issue of sustainability. As such, it would be difficult for any single country to conduct research on all of the fields related to sustainability science. Probably, each country has its own focused areas relative to the other fields. To see which fields are emphasized in the research of each country, the degree of concentration on a specific field is calculated as follows: cif ¼ nif ni Nf N ; where cif is the degree of concentration of country i on subject field f, nif is the number of articles of country i in field f, ni is the total number of articles of country i, and Nf is the number of articles of all countries in field f. Just as each country would have specialization with regard to research fields, collaboration between two countries could also be specialized in certain fields. In a similar manner, we examine the degree of research concentration on a specific field in international collaboration as follows: cijf ¼ nijf nij Nf N where cijf is the degree of concentration in international collaboration between country i and country j on subject field f, nijf is the number of co-authored articles between country i and country j in field f, and nij is the total number of coauthored articles between country i and country j. Results By using the methodology mentioned in the previous section, we collected a total of 41,487 articles on sustainability. The journals in which these articles were published are listed in Table 1. In this table, only those journals which published more than 100 articles on sustainability are included. Ecological Economics, Forest Ecology and Management, Water Science and Technology, and Agriculture Ecosystems & Environment are among the journals which published a large number of articles on sustainability. The trends in the number of countries which published scientific articles on sustainability are shown in Fig. 1. Countries have been increasingly engaged in publishing scientific articles on sustainability since the beginning of the 1990s. Along with that trend, the number of countries which have been involved in international collaboration in research on sustainability has also increased steadily. Approximately 90% of those countries with publications on sustainability have been engaged in some form of international collaboration, which suggests that, in most of the countries, there are researchers who are working on issues related to sustainability with researchers in different countries. A comparison is made between the papers written through international collaboration and those written domestically in Fig. 2. The proportion of publications through two-country collaboration is increasing, with the ratio of publications through bilateral collaboration to domestic research reaching more than one to five. The trend in publications through collaboration among three countries also shows a gradual increase compared with papers written without any international collaboration. To see the diversity of the countries which have been engaged in publishing scientific papers on sustainability, we examined the degree of concentration of the countries with publications. The result is shown in Fig. 3. The Hirschman–Herfindahl index (HHI) is adopted to examine the degree of concentration of the countries engaged in publishing scientific articles on sustainability. A large value of HHI means that the publication of scientific articles on sustainability is concentrated in a relatively small number of countries. While the number of scientific articles on sustainability has been increasing rapidly since the early 1990s, HHI for the countries which produce these publications has been declining, which suggests that more and more countries have been engaged in research activities on sustainability issues. HHI for the countries which have publications through international collaboration shows a trend of more significant decline, indicating that more diverse countries have recently been involved in international research collaboration. The geographical distribution of scientific articles on sustainability accumulated by 2007 is given in Fig. 4. Approximately one-fourth of articles related to sustainability were published by researchers who belong to organizations based in the US, followed by other Englishspeaking countries, namely, the UK, Canada, and Australia. Researchers in European countries, including Germany, the Netherlands, France, and Sweden, also published many articles, as well as those in Asian countries, notably, China and Japan. The network structures of international research collaboration on sustainability in 1995, 2000, and 2007 are shown in Figs. 5, 6, and 7, respectively. Since there are 177 countries which published at least one article related to sustainability, the network structure of each contains only the top 30 countries with regard to the number of publications in the respective year. The width of a link represents the collaboration-intensity similarity between the two countries connected by the link. To see more details on the network structure of international research collaboration on sustainability in 2007, Table 2 gives the list of country pairs with a large value of collaboration-intensity similarity. The list of collaboration pairs with a larger similarity value include countries which are located in geographical proximity, notably, Finland, Norway, Sweden, Denmark, Austria, and Italy in Europe. In other regions, Asian countries, namely, India, China, Thailand, Taiwan, and Japan, are also focusing on collaboration with each other. While the largest number of scientific articles on sustainability has been published by researchers in the US, the top twenty list of the country pairs with a large similarity value does not include this country, which suggests that authors in the US tend to collaborate relatively equally with their counterparts in other countries. Reflecting such close relationships among countries, we can see some groups who have dense ties within the group, as shown in Fig. 8. As can been seen, the collaboration network is basically fragmented into three regional blocks, namely, Europe and Africa, North and South Americas, and the Asia–Pacific region. This implies that the creation, transmission, and sharing of information and knowledge on sustainability might be limited within the regional clusters. The focused fields of research in sustainability science are given in Table 3 for some of the countries with a large number of publications on sustainability. The results show that the fields which have been emphasized in research activities with regard to sustainability are quite diverse between different countries. The values of the degree of concentration of focused fields in the US fall between -1.7 and 2%, and those values range from -1.5 to 2.3% in the case of China. This suggests that the research activities conducted in these countries are distributed evenly in different fields relative to the global trends. In other countries like Germany, India, and Japan, on the other hand, the range of values of the degree of concentration in subject fields is relatively large, which implies that research activities in these countries are focused on specific fields. In a similar way to the focused fields of each country, we examined the research fields which have been emphasized in international collaborations on sustainability. Table 4 shows examples of the focused areas in research collaboration between Japan and China. The field of water resources is the subject which is most frequently addressed in research collaboration between the two countries. While this area is also strongly emphasized in China, it is not the field which is of particular focus in Japan. This can also be said for other areas of strong focus of the bilateral collaboration, including plant sciences, ecology, soil science, and environmental sciences. These results suggest that the fields of research collaboration between Japan and China in sustainability science mainly reflect those areas emphasized in China. It would be implied that the bilateral collaboration is mainly influenced by the research agenda which includes urgent needs in China. While the necessity of environmental protection has been increasingly recognized and addressed in China, Japan has a substantial amount of research findings and expertise in diverse areas of environmental issues, partly due to the ample experiences of coping with various types of pollution and accidents accumulated in the past. Research collaboration between Japan and China, thus, would tend to be mainly aimed at addressing research needs in China, with the support of knowledge and expertise coming from Japanese researchers. Finally, we would like to address some of the limitations of this research. As we discussed in the introduction, we limited the scope of our examination to articles published in scientific journals by collecting those articles which include ‘sustainable’ or ‘sustainability’ in the title, abstract, or keywords. As extensive debates currently continue among researchers with diverse backgrounds with regard to the definitions, concepts, and methodologies in sustainability science, this is meant to be a first attempt to examine quantitatively the overall patterns of collaboration in the merging field. While this focus on quantitative analysis is very useful in dealing with a large number of academic articles published over a relatively long period of time, it could result in ignoring qualitative aspects which would be important in grasping a more in-depth understanding of the nature of this field. For example, public participation (Kasemir et al. 2003) and social learning (Ta`bara and

Pahl-Wostl 2007; Pahl-Wostl et al. 2008) have been emphasized in the previous literature as crucial aspects of sustainability science. One way to proceed to the next step is to focus on only articles that meet one or several of criteria discussed in sustainability science, such as interdisciplinarity and problem-focused approaches, and to conduct a qualitative analysis with explicit focus on contents. By so doing, we expect that the inter-disciplinary, problem-oriented nature of sustainability science will be explored in more detail.

Conclusion

As a preliminary step to conduct an institutional analysis of the emerging field of sustainability science, this paper examined quantitatively the patterns of research collaboration in fields related to sustainability over national boundaries. The results indicate that an increasing number of countries are engaged in research on sustainability, with the proportion of articles published through international collaboration rising as well. The number of countries engaged in international collaboration on sustainability research has been increasing, and the diversity of countries engaged in research collaboration beyond national borders is also increasing. The geographical patterns of collaboration on sustainability show that research collaboration tends to be conducted between countries which are geographically located closely. This suggests that communication and information exchange might be limited within the regional clusters. It is also shown that the focused fields of research activities on sustainability are significantly different between countries, as each country has its focused fields of research related to sustainability. The specialization of research activities is also observed in international collaboration. On the one hand, these patterns of international collaboration within regional clusters focusing on specific fields could be effective in promoting the creation, transmission, and sharing of knowledge on sustainability utilizing the already existing regional networks. The predominance of regional clusters, on the other hand, might pose a serious obstacle to collecting, exchanging, and integrating diverse types of knowledge, especially when it is necessary to deal with problems involving large-scale complex interactions with long-term implications, such as climate change. In that case it, would be of critical importance to establish inter-regional linkages at the global level. Several attempts have already been started to set up global schemes for research collaboration on sustainability science. Among the new types of organizational and institutional arrangements are the Alliance for Global Sustainability (AGS), an inter-university research collaboration between the University of Tokyo in Japan, Massachusetts Institute of Technology (MIT) in the US, Federal Institute of Technology (ETH) in Switzerland, and Chalmers University of Technology in Sweden, and the Forum on Science and Innovation for Sustainable Development hosted by the American Association for the Advancement of Science (AAAS). The International Conference on Sustainability Science (ICSS) has also been established to encourage communications and knowledge exchange on diverse issues linked to sustainability science. These emerging organizational and institutional arrangements will have significant implications for the future development of sustainability science, which is chiefly aimed at tackling the challenge of producing, communicating, and integrating diverse types of knowledge and expertise for the sustainability of the Earth.

Open research data increasingly receives attention from research funders, governments, and academic institutions (e.g. European Commission, 2019; European Research Council Executive Agency, 2017; OECD, 2015; Patel, 2016; Saywell and Crocker, 2019; Ventura et al., 2020; Zuiderwijk and Spiers, 2019). In this paper, we define open research data as research data that is freely and publicly made available for the long term, in open, interoperable, and machine-readable formats, accompanied by sufficient metadata, and legally fit to be crawled, reused, and modified (Australian National Data Service, n.d.; Austin et al., 2017; European Commission, 2017; Kondo et al., 2018; Lindman and Tammisto, 2011; Open Knowledge Foundation, n.d.; Zuiderwijk and Spiers, 2019). Open research data includes quantitative and qualitative data in various forms and types, including observational, experimental, theoretical, and computational data. Openly sharing research data is expected to increase the cost-effectiveness and democratization of the data (Borgerud and Borglund, 2020). Another potential benefit is that access to research data allows for the verification of results and reduces the duplication of research efforts, both inside and outside academia (Patel, 2016; Saywell and Crocker, 2019; Zuiderwijk and Spiers, 2019). Open research data can also increase researchers’ citation count and acknowledgment of their work (Bullini Orlandi et al., 2019; Fecher and Friesike, 2014; Fries, 2014; Mosconi et al., 2019; Piwowar et al., 2007; Steel et al., 2019; Yarime, 2017). However, often for good reasons, researchers are reluctant to share their research data openly and to reuse data shared openly by other researchers. For example, researchers may be lacking the time, money, and necessary data management skills to share and reuse research data (European Commission, 2019; Gertrudis-Casado et al., 2016; OECD, 2015; Yarime, 2017; Zuiderwijk et al., 2012). Another inhibitor is a perceived sense of competition with other researchers: researchers may fear that other researchers scrutinize their data and discover something novel before they do (Borgerud and Borglund, 2020; Yarime, 2017). Moreover, researchers may be reluctant to share and reuse research data due to legal matters such as copyright, licenses, and data (privacy) sensitivity (Borgerud and Borglund, 2020; Patel, 2016; Viseur, 2015). The majority of challenges for open research data sharing and use cannot be mitigated completely. Nevertheless, the negative impact of many challenges can be reduced with suitable infrastructural and institutional arrangements, as suggested by previous research (Altayar, 2018; Zuiderwijk, 2015). In this study, we refer to arrangements as the combination of individual instruments. We define institutional arrangements as the combination of instruments related to formal structures (e.g. university policies), informal structures (e.g. norms, culture), and operational mechanisms (e.g. existing data-sharing processes) that research institutions can employ to incentivize open research data sharing and use (derived from North, 2005; Williamson, 2009). Examples of institutional instruments include implementing data-sharing policies (Patel, 2016) and offering educational programs about research data management (Kondo et al., 2018; Steel et al., 2019). We define infrastructural arrangements as the combination of instruments related to technical elements (e.g. open data portals, (meta)data standards and formats, and tools for processing, searching, analyzing, and visualizing data) and governance elements (e.g. mechanisms to enhance privacy, trust, and interaction with other data providers and users) to stimulate open research data sharing and use (derived from Zuiderwijk, 2015). Examples of infrastructural instruments include offering openly available infrastructures (Patel, 2016) and having data quality indicators on the platform (Charalabidis et al., 2014). The term open data infrastructure may refer to multiple data sharing and reuse environments, such as repositories, archives, portals, and platforms. These infrastructures must account for a wide range of data sources and (meta) data semantics (Abbà et al., 2015), which may vary per domain (Borgerud and Borglund, 2020; Neuroth et al., 2013). Infrastructural and institutional arrangements can be studied in isolation, yet combining them is expected to increase their effectiveness. For example, the challenge of “lacking rewards for sharing open research data” may be addressed by a scientific output assessment system that adequately takes account of open data sharing contributions of researchers (European Members of the International Council for Science, 2018) combined with acknowledging researchers’ data sharing behavior on their institutional website (infrastructural and institutional instrument). As another example, the challenge of “lacking skills to use open research data” may be addressed by improving the ease of use of open research data portals (infrastructural instrument) in combination with researcher training for using such portals (institutional instrument). While previous research provides an overview of possible institutional and infrastructural instruments (e.g. Fecher and Friesike, 2014; Patel, 2016), there is a lack of insight into what combinations of instruments (i.e. arrangements) positively affect research data sharing and reuse in particular contexts. Some arrangements may be more valuable in certain situations than others. For example, infrastructural and institutional arrangements need to be adjusted to the knowledge and skills of the involved researchers and support staff. Moreover, the country, organizational structure, and information needs of an institution shape its information and knowledge management systems (Deja, 2019). Due to the lack of deep, contextual insight, it is currently difficult to assess the effectiveness of various arrangements for different situations. This research aims to investigate which combination of institutional and infrastructural arrangements have a positive impact on research data sharing and reuse in a specific situation, using a case study research approach. Our contribution to the library and information science literature is twofold. First, this study is among the first to provide a contextualized overview of institutional and infrastructural arrangements potentially useful for universities and academic libraries to stimulate open research data sharing and reuse. Second, this study discusses the potential impact of implementing certain arrangements in the particular case of a Dutch university active in open science policy implementation. Practically, at university and academic library level, the lessons learnt from this study may be useful for open research data policymakers, staff supporting research data infrastructure development and use, and individual researchers. Ultimately, our findings should allow researchers to examine the effectiveness of the identified arrangements in different contexts. Finally, implementing adequate infrastructural and institutional arrangements enables research institutions and research data infrastructure providers to better support and stimulate data sharing and reuse behavior by researchers. This paper is structured as follows. First, we provide an overview of infrastructural and institutional instruments as identified from previous research. Next, we describe the case study research approach adopted in this study, and discuss the contextual aspects that should be kept in mind when interpreting our case study findings. Thereafter, we present the case study findings and discuss the implications of our findings for university policy makers (including research support staff), governmental policymakers, research funding agencies, and individual researchers. Finally, we present our research conclusions.

Research background

This section provides an overview of infrastructural instruments (section “Infrastructural instruments to support open research data sharing and reuse”) and institutional instruments (section “Institutional instruments to support open research data sharing and reuse”) to support research data sharing and reuse as derived from previous research. We identified these instruments through a literature review using the queries “(open research data) AND (infrastructural OR institutional) AND (arrangement\*)” and “(‘open research data’) AND (infrastructure\* OR institution\*)” in Scopus. The systematic literature review was complemented with a snowballing technique, as recommended by Jalali and Wohlin (2012). The relevance of all results was first assessed based on the title and abstract. The remaining results were further judged on their quality, actuality, and relevance.

Infrastructural instruments to support open research data sharing and reuse

This section describes instruments that infrastructure providers can use to support data sharing and reuse. Table 1 summarizes the infrastructural instruments derived from previous research. The first group of infrastructural instruments identified from the literature concerns improving the usability of open research data infrastructures. The availability of research data infrastructures with sufficient user functionality is considered a fundamental driver for open data use (European Commission, 2018; OECD, 2015). Infrastructures should be easy to use and comprehensive, as they should cover data provisioning, searching, downloading, and processing options (Charalabidis et al., 2014). Currently, researchers encounter a lack of common nomenclature and (meta)data standards across different infrastructures (Borgerud and Borglund, 2020). Moreover, Patel (2016) underlines the importance of presenting data accurately to prevent incorrect interpretation of the data by others. In the same vein, open data infrastructures should offer communication functions and data quality indicators to introduce feedback loops between data authors and users (Charalabidis et al., 2014). The second group of instruments focusses on offering assistance with legal aspects of sharing and reusing open research data. For instance, research may be unsure of how to ensure the privacy of respondents in their data (Borgerud and Borglund, 2020). Viseur (2015) and Patel (2016) argue that researchers should be offered tools to help them choose and apply a license to their work, to take care of their copyright and intellectual property concerns. A third group of instruments deals with assisting researchers in choosing an appropriate infrastructure for sharing or searching open research data. A first way to achieve this is listing data repositories in repository aggregating registries such as Re3Data.org, making infrastructures themselves more findable (Francis and Das, 2019). Applying this FAIR-principle on a meta-level improves value for everyone involved due to network effects: the aggregator users have more repositories to search through, increasing the odds they find fitting data. Subsequently, the aggregator, the enlisted data infrastructures, and the corresponding data publishers see their audience grow. A second method to make repositories more findable is by integrating infrastructures in other technological research environments. Bullini Orlandi et al. (2019) find that ICT tools enabling sharing of scholarly work are less likely to be adopted by researchers than tools supporting the actual research phase. Steinbeck et al. (2020) see the integration of data repositories with other research tools (e.g. data viewers, editors, processors, and search engines) as fundamental for supporting researchers in all stages of the research workflow. The fourth group of instruments is about ensuring the sustainability of open research data infrastructures. This is firstly a technical matter, as data must be stored on storage media and in formats fit for long-term preservation (Borgerud and Borglund, 2020; Mosconi et al., 2019). Patel (2016) also calls for data security to be guaranteed. Secondly, long-term funding must be secured (OECD, 2015; Zielinski et al., 2019). Kitchin et al. (2015) note that securing funding is especially challenging for non-national open data infrastructures, as they are less likely to receive sufficient state subsidies to compensate for their costs. The infrastructural instruments described above can be implemented by research institutions to support and stimulate open research data sharing and use. They can be combined with the institutional instruments as mentioned in the next section. Institutional instruments to support open research data sharing and reuse

Table 2 provides an overview of institutional instruments available to research institutions (such as universities) to support data sharing and reuse by their researchers. The first institutional instrument category identified in the literature concerns offering researchers sufficient credit for sharing and reusing open research data. Steel et al. (2019) state that scholars can be provided tangible valueaddition by increasing their knowledge and management skills of the research life cycle. This maximizes the impact of their data publications and hence increases their acknowledgment. Although publishing datasets is usually less rewarding than publishing articles (Bullini Orlandi et al., 2019; Mosconi et al., 2019; Steel et al., 2019), data publications may still yield additional recognition. They enhance the visibility of corresponding articles (Piwowar et al., 2007), and the average number of citations influences the reputation and funding prospects of researchers (Fecher and Friesike, 2014). Mack (2019) advocates for changing the mindset of personnel and support accordingly. Codifying data sharing requirements in a policy aids this culture change, as it demonstrates the dedication of the institution toward these ambitions (Patel, 2016). Institutions should also acknowledge researchers’ data sharing efforts better at an institutional level. For example, institutions can evaluate research output in terms of shared data (Yarime, 2017; Zuiderwijk and Spiers, 2019). For this, a scientific output assessment system such as “altmetrics” can be used (European Members of the International Council for Science, 2018). The second instrument category intends to improve the data management skills of researchers. Researchers must possess the necessary research data management skills to publish and reuse data (OECD, 2015; Zuiderwijk et al., 2012). Vital competencies include data searching, selection, processing, analysis, and presentation (Gertrudis- Casado et al., 2016). Data literacy enhancement programs by librarians can improve data management skills (Kondo et al., 2018; Steel et al., 2019). However, it is unsure how appealing this is to researchers: following trainings requires a time investment and, consequently, a change of formerly established routines. Nonetheless, it is likely that researchers are willing to invest some time in training, given increasing pressure by funders to share research data after research completion (European Commission, 2019; European Research Council Executive Agency, 2017; OECD, 2015; Saywell and Crocker, 2019; Ventura et al., 2020). The third institutional instrument category concerns offering assistance with legal aspects of sharing and reusing data. The ability to deal with legal aspects of both sharing and reusing research data is a subset of the aforementioned data management skills. Besides education through courses, Mack (2019) recommends the availability of on-site experts who researchers can consult for legal issues. Patel (2016) also advocates for dealing with copyright issues, data licensing issues, privacy issues, and data protection issues on an institutional level. She suggests the development of copyright and privacy statements. As section “Infrastructural instruments to support open research data sharing and reuse” highlighted, dealing with these legal issues can also be done effectively on an infrastructural level. It is plausible that using these instruments in tandem yields the best results, as they complement each other. The fourth institutional instrument category entails covering costs that open research data sharing and reuse incur. Even if researchers are willing and able to share data, they may require financial support to do so (European Commission, 2019; Mack, 2019; OECD, 2015). Some funders reimburse these data-sharing costs (OpenAIRE, 2018). Reusing data comes with costs as well (OECD, 2015); for instance because researchers must examine whether data they find fits their research and because they may have to transform the data before using it. Institutions could offer financial support to researchers not compensated by their funders. A fifth category is taking research data management tasks out of researchers’ hands. Scientists could hire a data manager to take care of research data management for their projects (Bishop et al., 2021). Their tasks may include writing data management plans, sharing data, and identifying public research data that can be reused for the project (Bishop et al., 2021). Data managers are thus valuable for researchers with insufficient time or data management knowledge (Tenopir et al., 2020).

Case study research approach

In this section, we describe the motivation for taking a case study research approach (section “Motivation for case study research approach”), and explain our study selection criteria (section “Case selection criteria”) and information sources (section “Case study information sources”). We finish by describing how we analyzed the interview transcripts (section “Analysis of interview transcripts”).

Motivation for case study research approach

We adopted a case study research approach to investigate which institutional and infrastructural arrangements positively impact research data sharing and reuse. The qualitative case study method “investigates a contemporary phenomenon (the “case”) in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident” (Yin, 2018: 45). Yin (2018) recommends the case study approach for situations in which “(1) your main research questions are “how” or “why” questions, (2) you have little or no control over behavioral events, and (3) your focus of study is a contemporary (as opposed to entirely historical) phenomenon” (p. 32). Moreover, case studies are suitable when personal experiences and the context of certain behavior play a fundamental role (Benbasat et al., 1987; Bonoma and Wong, 1985). The above-mentioned aspects characterize this research. First, this study is highly exploratory: the phenomenon under investigation has not been studied extensively yet. Previous research does not address the impact of contextual aspects on the effectiveness of institutional and infrastructural arrangements to support open research data sharing and reuse. Second, we aim to provide a deep and rich understanding of how and why specific arrangements positively impact open research data sharing and reuse in a particular context, in this way answering “how” and “why” questions. For this reason, the authors adopted an interpretative approach. It is not the goal to formulate an objective truth, but instead to represent the perspectives and interpretations given in the case study information sources (Walsham 1995). The inferences and conclusions that result from these inductive analyses can then be extrapolated to other cases (e.g. universities or academic libraries) (Walsham and Waema, 1994). Third, we, as investigators, have no control over the data sharing and reuse behavior. And fourth, this research examines current data sharing and reuse behavior, rather than historical phenomena.

Case selection criteria

In case study research, the researchers typically define case selection criteria and then select a specific case (Yin, 2018). Using theoretical sampling, our case study selection criteria were as follows: 1. The case concerns research data sharing and reuse in the Netherlands and enables access to relevant interviewees. Since the authors were located in the Netherlands and speak the language, a Dutch case would be most convenient. 2. In the case, both infrastructural and institutional arrangements are implemented. 3. The case concerns a Dutch public university, as we assume that non-public and commercial research institutions have less incentive to stimulate sharing research data. 4. The university must have extensive experience in stimulating research data sharing and reuse. As this experience is hard to measure, we defined two proxies. Out of all Dutch public universities, we selected universities that a) have a research data management policy available on the internet, and b) supplement this with publicly available facultyspecific policies for the majority of their faculties. At the moment of our study, four Dutch public universities fulfilled these criteria: Delft University of Technology (TU Delft), University of Groningen, Utrecht University, and Free University Amsterdam.1 For practical reasons, primarily easy access to interviewees, we selected the TU Delft as the institution to study. Our selected case thus concerns the institutional and infrastructural arrangements that the TU Delft implements to stimulate data sharing and reuse, and what experience researchers, policy makers, and support staff have with these arrangements. Section “Case description: Research data sharing and reuse at Delft University of Technology” describes the selected case in more detail.

Case study information sources

To reduce bias in case study research, Yin (2018) recommends the use of multiple information sources. Table 3 provides an overview of the information sources for our case study. Table 4 describes the roles of the selected interviewees and their experience with research data management. The interviewed data steward and researchers (i.e. I3, I5, I6 and I7) are all employed at the same faculty (which we do not mention for confidentiality reasons). The other interviewees (i.e. I1, I2, and I4) operate on a university-wide level.

Analysis of interview transcripts

The interviews covered five parts: (1) the interviewees’ background, (2) the (perceived) familiarity of researchers with the support options for openly sharing and reusing research data available at the institution, and the demand for these options, (3) why they believe researchers seek support for research data sharing and reuse, (4) the (perceived) familiarity, satisfaction, level of understanding, and compliance by researchers with research data management policies and resources, and (5) how the interviewees think researchers use the available research data infrastructures to upload or find research data. We gave interviewees time to elaborate on personal experiences and their motivations or struggles when sharing or reusing research data. We fully transcribed and anonymized the interviews. Interviewees could review the transcripts, which resulted in adding two minor clarifications. Next, we coded the transcripts using the ATLAS.ti version 9.1.7 software, a standard tool for qualitative document analysis (Alasseri et al., 2018). We applied theory-driven coding to the transcripts, meaning that we developed labels to make sense of the interview data based on theory and concepts before starting the coding process (DeCuir-Gunby et al., 2011). For this step, we used the literature described in section “Research background.” As recommended by DeCuir- Gunby et al. (2011), we then applied open, axial and focused coding. For the open coding process, we followed the recommendations by Lindlof (1995) and started the open coding process by reading every transcript. We then marked and categorized relevant parts and built codes around the identified categories. This process resulted in a long list of codes. Next, we reviewed the codes in their context using axial coding. Axial coding let us relate (sub)categories and their properties, allowing us to make the “disassembled” pieces of the transcripts into one coherent analysis again (Charmaz, 2006; Corbin and Strauss, 1988). During this step, we put the codes into overarching categories. Finally, we tested the reliability of the resulting codes using focused coding. Focused coding entails “using the most significant and/or frequent earlier codes to sift through large amounts of data” (Charmaz, 2006: 57). We identified the most important variables, reread the transcripts, and selectively recoded the transcripts where applicable. In this coding round, we also split certain codes to apply to only researchers (indicated with the suffix [R]) or to policymakers and support staff (suffix [PM/SS]). The coded, anonymized transcripts and codebook are publicly available as supplementary data at https://doi.org/10.4121/19635147.

Case description: Research data sharing and reuse at Delft University of Technology

This section describes the case we studied, namely the infrastructural and institutional arrangements for research data sharing and reuse by Delft University of Technology. The TU Delft is a Dutch university that is the scholarly home to roughly 27,300 students, 6300 staff members, and 2900 Ph.D. candidates (TU Delft, 2022). Its road to Open Science is long: the first initiatives of the university started almost two decades ago (TU Delft, 2012). The institution signed the Berlin Declaration on Open Access in 2005 (TU Delft, 2016). The university considered open access a strategic priority in its Strategic Plan for 2012 to 2017 (TU Delft, 2012) and enacted a separate policy on Open Access Publishing in 2016 (TU Delft, 2016). In the Strategic Framework for 2018–2024, this focus was broadened to Open Science in general (TU Delft, 2018). In 2019, the university fostered a Strategic Plan Open Science for 2020–2024 (Haslinger, 2019), which included a policy on research software sharing (Akhmerov et al., 2021). More specifically, for research data management, the university set up institutional arrangements in 2017. The university defines successful research data management in terms of three fundamental pillars: infrastructure, culture, and policy (Dunning, 2018b). Hence, the university introduced data sharing funding opportunities and educational resources, and installed data stewards and research data officers at the university library and all faculties (Dunning, 2017). The Research Data Framework Policy followed in 2018 and was updated in 2020 (Ahlers et al., 2020). This document forms an overarching framework for the entire university. The framework is supplemented by facultyspecific policies, first developed in 2018 (Dunning, 2018a). One critical instrument, that has been in place since 2010, is the 4TU.ResearchData Repository. The TU Delft manages this repository with Eindhoven University of Technology, the University of Twente, and Wageningen University & Research, all Dutch universities. Despite these institutional ties, the infrastructure offers its services to researchers worldwide (4TU.ResearchData Consortium, n.d.-c). In the year 2021, 835 datasets were published on the repository (4TU.ResearchData Consortium, n.d.-d). In the same year, all datasets uploaded amassed over one million views and 330.000 downloads (4TU.ResearchData Consortium, 2022).

Case study findings: Infrastructural and institutional instruments in practice

Our case study revealed how infrastructural and institutional arrangements for data sharing and reuse are perceived in practice. Below, we categorize the identified instruments using the categories derived in section “Infrastructural instruments to support open research data sharing and reuse.” The categories are underlined and the identified instruments are depicted in italics. When we refer to statements of the interviewees, we refer to the interviewee number between square brackets. Infrastructural instruments

Table 5 categorizes the identified infrastructural instruments. First, we identified infrastructural instruments that influence the usability of infrastructures for openly sharing and reusing data. We found that researchers are more inclined to use a data sharing infrastructure if they are offered support for the infrastructure compared to when this support lacks. [I1] provided an example of the support offered by the 4TU infrastructure: “When somebody in the back-office sees [. . .], maybe, the lack of certain information, certain documentation, or maybe they think “okay, maybe this license might not be the appropriate one”, they can contact the researchers.” However, this back-office review is also time-consuming, whereas most researchers want to publish quickly [I2]. Researchers choosing to publish quickly miss out on the added value of the metadata check and findability and usability of the published data may then decrease. Repositories with a quicker data sharing process might become contaminated with poorer datasets, requiring more effort from the data searcher to assess their usefulness [I1]. Moreover, several interviewees named user convenience as a factor determining research data usage [I2, I4, I7]. In practice, most infrastructure providers understand that convenience can make or break infrastructure usage; they thus have their usability well in order [I2]. In addition, researchers are increasingly medialiterate and thus capable of navigating infrastructures [I1]. Related to user convenience is the expressed desire for infrastructures to facilitate multiple publications of a single dataset through version control [I5] and anonymous data publication for the peer review of a related journal paper [I5]. A less technical instrument concerns guidelines on when sharing data is recommended [I5]. Also, some researchers prefer open-source infrastructures over infrastructures provided by (commercial) parties with a reputation of not supporting open science [I2]. The most important instrument related to legal aspects concerns offering support for choosing licenses when uploading data. Licenses are useful, as they clarify how data can legally be reused. Researchers often struggle with identifying an appropriate license and should be informed better [I1]. License information can be provided on both an institutional and infrastructural level. We derived three instruments to assist researchers in choosing an appropriate infrastructure for sharing or searching open data. Regarding infrastructure identification, the interviewees mention the lack of a powerful search engine for finding relevant open datasets. [I3] explained the problem as twofold: on the one hand, it is hard to write queries that match a specific research question. On the other hand, finding relevant data requires datasets to be carefully described with formal language and metadata. [I3] continued: “[t]he connections between the datasets, right now they’re established based on text comparison. I did a search on 4TU or Zenodo [. . .] about the medieval potato crisis [. . .] and I ended up with poetry and all kinds of things that were completely unrelated. We don’t really know yet how to efficiently search and connect datasets to discover them.” The interviewed researchers indicated that having a powerful search engine would make them more likely to search for data: even if they do not expect to find data, the lowered transaction costs make it a small effort in exchange for potentially large benefits [I6]. Cataloging known infrastructures could also help researchers searching for an appropriate infrastructure. This can be done both by linking infrastructures to infrastructure aggregators and by spreading knowledge about (domain-specific) infrastructures in research groups. Finally, the interviews clarified how the integration of data infrastructures in other research tools lowers the bar to use these infrastructures. This applies to both up and downloading research data. For example, [I6] said: “If everything is concentrated into one interface, for example, Google Scholar, I would definitely take a look because it would reduce the overhead.” This integration seems, however, hard to accomplish. It requires the cooperation of (mostly) commercial parties. An easier alternative may be to connect these infrastructures to digital tools offered by research institutions, such as, an intranet, university workstations, or SURFDrive (an academic counterpart of cloud storage services).

Institutional instruments The previous section discussed our case study findings regarding infrastructural instruments. This section presents the findings concerning institutional instruments for open research data sharing and reuse (see Table 6). Interviewees spoke about various instruments related to offering researchers credit for sharing and reusing open research data. First, when starting with institutionalizing research data management, one may be inclined to write up a policy and then shape the other institutional instruments accordingly. However, all interviewed policymakers recommended starting by doing: putting practice first and policy second. They prefer installing support staff and establishing a community first. The data management policy documents can then be shaped accordingly. [I1], [I2] and [I4] recommend to implement a generic universitywide framework and supplement this with faculty-specific policies, “because workflows depend or differ a lot by the type of research they do; at different faculties there will be different workflows” [I1]. Moreover, several interviewees state that the data management policies should not be used as coercive instruments [I1, I2, I4]. They should instead serve as a measure of last resort in case of unwilling researchers, given the administrative burden that policies impose on researchers [I1, I2, I4]. All policymakers and support staff deem changing the mindset and advocating for proper data management from the bottom up more effective. [I2] even claims that “[t]he policy is the least important thing. The best way to change people’s minds is with people like the data stewards and the data managers. So the informal support, right, just there next door.” The interviewees also referred to the institutional instrument “Data Management Plans (DMPs).” DMPs force researchers to think about how they will manage and publish their data, and may therefore be required by funders and ethics committees. This is good to get researchers involved with the subject [I2, I3, I4]. However, similar to policies, DMPs are usually perceived as an obligation and met by equivalent resistance [I1, I2]. Researchers may cut corners to just “tick the box,” as [I3] describes: “For project proposals, they don’t care so much. They just want to have the document so that their proposal is complete, [. . .] they are less worried about will the data be properly stored” [I3]. Furthermore, one policymaker aptly noted that “[f]illing a DMP does not really mean you will change your habits when you are doing research. That is much more difficult to measure” [I2]. An extra complicating matter is that institutions and funders may enforce inconsistent policies and DMP templates. These should therefore be standardized within domains, both nationally and internationally [I4]. A positive illustration is that the Dutch Research Council (NWO) has its own template, but also approves of the TU Delft DMP template, so that researchers can comply with requirements of both parties with a single DMP (Clare et al., 2021). Currently, the university library aims to make data management plans “less of a form and more of a service” by taking steps toward machine-actionable DMPs. Particular information entered in a data management plan is then processed and linked to different services automatically [I1, I2]. For example, once a question about the size of the data to be produced in the research is answered, a message is sent to the information technology department to allocate storage to the researcher [I1]. Another example is that if the data management plan contains dates, the researcher receives an automated reminder to either take the corresponding action or update the plan when the date is nearing [I2]. In the second category of institutional instruments, all interviewed policymakers and support staff talked extensively about how they aim to improve the data management skills of researchers. In the examined case, the university offers three educational programs on research data management and open science. First, a relatively generic Open Science Massive Open Online Course (MOOC) teaches learners about open science in general and a bit of data management (de Jong et al., n.d.). Second, the university library hosts a research data management course to new PhD-students and other interested researchers. This course teaches data management fundamentals and the support arrangements in place. Third, the university library offers three research data management workshops: the Software Carpentry, Data Carpentry, and Code Refinery workshop. Besides educational programs, the university library maintains research data management guides on the university website. The impact of these guides is low; [I2] is not too enthusiastic about them: “I think putting lots of information on the library website doesn’t work so well, because then people have to navigate the pages, they have to understand the language, it takes time.” This is confirmed by [I7], who had consulted a guide on research data management on the institutional website and found a figure indicating what steps to take in data management, which was simply too complicated (TU Delft Library, n.d.-b). Furthermore, in the examined case, spreading knowledge through the researcher community was deemed a powerful institutional instrument. One way to do this is through the Open Science Community Delft; a group of data management enthusiasts that researchers can approach for any data queries (Clare, 2019). Moreover, the university launched the 4TU.ResearchData Community last year, as well as working groups on FAIR code and privacy issues (Clare et al., 2021). The installation of data stewards in faculties also fits this arrangement. Likewise, the interviewed policymakers agreed that communication about research data management (developments) should happen through community and peer-to-peer contact or by educating new researchers. [I4] underlines the importance of peer behavior: “If you see that people in your research groups or research area, nationally or internationally, are doing that, then you probably will follow that behavior. So you need some good champions, people who are known in the field, for making that first move.” Moreover, all interviewees state that appointing a Data Steward in each university faculty has proven its value extensively. The library is responsible for coordinating the network of data stewards (Ahlers et al., 2020). Data stewards are the first contact point for any data-related query for anyone in the faculty. Because of their research background, researchers perceive them more as peers than as administration [I1]. The stewards also approve of and provide feedback on data management plans, so every researcher that fills one in knows of the possibility to ask them questions [I1, I3]. The interviewed data steward [I3] indicated that more than half of the researchers approach him because they need approval on a document. He said: “The fun thing is that despite being formal things that they have to do, I get rather positive feedback after the meetings that we have once we’re done working with the DMP. Usually, people say ‘oh, well that was actually helpful.’” All interviewees highly appreciate the informal nature of the contact with data stewards. Nonetheless, senior researchers are used to the availability of formal support services and tend to go for those, as they operate more independent than, for example, PhD-students [I3]. Finally, for institutional instruments to achieve their value, researchers must be aware of their presence. Data stewards as the primary data management contact points for researchers can point researchers to other support instruments [I1, I4]. In addition, several educational instruments could be used to raise participants’ awareness about arrangements. Leaders of large research groups following data management courses can spread knowledge of arrangements and good practices among their peers. Moreover, [I4] notes that awareness of specific arrangements should be raised at tactical moments in which the researcher may need them, for example, when a new employee or research project starts [I4]. In the category of covering costs that open research data sharing and reuse incurs, the interviewees referred to the instrument of offering funds for refining and publishing research data openly. The examined university offers researchers funding to make their data shareable when they are (nearly) done with their corresponding article, or publish it even if no corresponding article is written. They can use the money to, for instance, hire a student assistant to properly anonymize and document their data [I1] (Clare, 2021). However, the funds have a minimal impact, as they do not always match researchers’ needs. For example, the funds usually come into play when a research project is nearing its deadline [I3]. Therefore, squeezing in all the work related to data sharing is problematic as the funds do not extend the projects themselves. Moreover, it is hard to estimate research data management costs [I3]. Finally, in the category of taking research data management tasks out of researchers’ hands, we identified the institutional instrument to appoint data managers to take over research data management tasks from researchers. For example, for research groups without the time or knowledge to take on data management tasks in their research projects, the university library employs two data managers who can be hired for projects’ data management tasks [I2]. They can be financed by the university library or, ideally, by the project funds [I2]. The impact of this instrument is not known yet, as the interviews preceded the installation of the data managers. We expect that the short-term effect of hiring data managers is positive, as they assist researchers in identifying fitting datasets for their research, and guarantee proper data sharing. However, the data managers’ availability also means that researchers lose an incentive to develop their own research data management capabilities, as they can simply delegate data management tasks to someone else.

Discussion

Our study suggests that combining infrastructural and institutional instruments in useful, contextualized arrangements can enhance their effectiveness. The interviews with, firstly, the researchers, and secondly, the policymakers and support staff made clear that researchers encounter different challenges in different parts of the research data management process. An important notion is that mitigating only a subset of these barriers will not suffice, as each of these barriers—such as lacking financial means, lacking knowledge on how data sharing or reuse, or not knowing which infrastructure to use—have the ability to block the data sharing or reuse in its entirety. Relatedly, applying the instruments complementarily increases their effectiveness. For example, appointing faculty data stewards will help less knowledgeable researchers to comply with formal data sharing policies. Similarly, educating researchers on how to share or reuse data provides a window for informing them about other institutional and infrastructural instruments available to them, so that they know where to look in which situation. Institutions and infrastructure providers should therefore maintain a broad perspective on the entire research data management process and combine different instruments to increase their individual effectiveness. In sum, the findings emphasize that instruments should be combined to form arrangements that tackle multiple data sharing and reuse challenges, on both an institutional and infrastructural level. Studying or implementing individual instruments insufficiently stimulates data sharing and reuse, as the experienced difficulties are intertwined and span a wide spectrum. Therefore, we recommend both practitioners and scholars to maintain a broader and overarching perspective, rather than focus efforts on one or a limited number of disconnected instruments. In interpreting this study’s findings, the reader should consider various contextual aspects that influence open research data sharing and reuse in our specific case study. First, the examined university operates in certain research disciplines and one of its key objectives is to create impact for a better society (TU Delft, n.d.-c, 2018). Because research at the TU Delft focuses on technical-scientific solutions to societal problems (TU Delft, n.d.-a, n.d.-b), the scientific knowledge developed can, indeed, easily be deployed for societal gain. Publishing new-found knowledge and corresponding research data is thus part of fulfilling this mission. In other words: the research disciplines in which the TU Delft operates, could be of more open nature than other disciplines, such as genomics (Piwowar and Vision, 2013), genetics and life sciences (Campbell and Bendavid, 2003), and astronomy (Wallis et al., 2013; Zuiderwijk and Spiers, 2019). The university acknowledges that some disciplines (e.g. “microscopy data, material science, the life sciences, hydrology”) have more momentum and stronger disciplinary communities discussing data sharing than others (Teperek et al., 2019). Another crucial contextual aspect is that the TU Delft has already made significant progress in the field of open science and open research data in the past years (see section “Case description: Research data sharing and reuse at Delft University of Technology”). Generally speaking, the research data management skills of its researchers are thus likely to be better developed than those of researchers in less experienced universities. Thus, researchers of the TU Delft need different support arrangements than researchers at institutions less familiar with open data, as changing research practices call for tailored support services (Cooper et al., 2019). For example, researchers that are only just getting into research data management are more likely to share a series of frequently occurring questions; they could be better off with generic educational arrangements to first increase their general data management skills. Furthermore, the TU Delft is located in the Netherlands and thus affected by the Dutch culture. The Dutch culture is strongly feminine and values inclusivity, solidarity and consensus, rather than competition and success (Hofstede Insights, 2021). Researchers sharing their data to facilitate research by others might thus well be a result of Dutch culture. The Netherlands also score relatively high on the scale of long-term orientation (Hofstede Insights, 2021), which results in a willingness to invest resources for dealing with future challenges. Investing in institutional and infrastructural arrangements, and consequently the actual sharing and reuse of research data, can be considered a necessity on the long term given benefits such as reducing the duplication of research efforts, and increasing the costeffectiveness and democratization of research. On the other hand, the Netherlands scores relatively high on the scale of individualism. Sharing data for the benefit of others seems out of line with this characteristic. However, data sharing and reuse also benefits the sharer (see section). It could be that researchers in the Netherlands are more strongly motivated by arrangements with a tangible benefit for themselves. Finally, we note that open science has long been on the Dutch political agenda. In 2013, the Dutch government took steps to accelerate the shift to open access publishing of articles, by bringing stakeholders together in both a national and international context (Dekker, 2013). The Netherlands further pushed for open science during the Dutch presidency of the European Union in 2016 (Enserink, 2016), hosting a conference that resulted in the “Amsterdam Call for Action on Open Science” (NLU, 2016). This document contains concrete actions to stimulate open access to publications and sharing of research data. In 2017, the National Plan Open Science was presented, wherein relevant stakeholders signed a “Declaration on Open Science” and constituted the National Platform Open Science (NLU, 2016). In this same context, a report on the Dutch data landscape was published in 2020, proposing actions to optimize sharing and reusing of research data (de Vries et al., 2020). These national policies, plans, and ambitions could explain the progress that TU Delft has made in the field of open science, resulting in different institutional and infrastructural arrangements compared to universities in other countries.

Conclusion This research aimed to investigate which combination of institutional and infrastructural arrangements positively impact research data sharing and reuse. Using a case study approach, we found that the institutional arrangements offered should at least include financial and administrative instruments (e.g. fulfilling formal requirements, such as funder or institutional policies). Moreover, institutions should strive for disciplinary standardization of policies and data management plans, to prevent hindering collaborations outside the university. Furthermore, our interviewees deemed changing the mindset of researchers fundamental: researchers should become aware of the benefits of sharing data. We found that universities recently starting with data sharing encouragement are best off installing non-policyrelated infrastructural and institutional arrangements before creating a policy. Finally, implementing legal instruments (e.g. handling licenses and privacy issues) and operational instruments (e.g. increasing data management skills and helping in identifying fitting infrastructures) are considered to positively impact research data sharing and reuse; these instruments can be implemented on both an infrastructural and institutional level. The interviewees lauded that data stewards function as an approachable one-stop-shop for support in all named support areas. This study contributes to the library, information science and open science literature by providing insight into how the instruments recommended by the literature should be adapted and applied to function properly in the context of a specific case. To the best of our knowledge, our study is the first to show how infrastructural and institutional arrangements for data sharing and reuse can be combined to increase their effectiveness. Moreover, this study revealed the importance of timing; researchers must be made aware of the existing infrastructural and institutional arrangements at the moments when they most likely need them. Using our framework of infrastructural and institutional instruments discussed in the literature and applied in our case study, other scholars can examine the effectiveness of the identified infrastructural and institutional instruments in other contexts and derive new insights on contextualization. As a societal and practical contribution, open data policymakers, universities, and open data infrastructure providers can use our findings to stimulate data sharing and reuse in practice, adapted to the contextual situation. This study focused on a single case and, besides several other information sources, it involved seven interviews with staff, particularly from the university library and a single university faculty. The findings may thus not necessarily reflect the situation of other faculties, institutions, and infrastructures. In addition, the authors of this paper work for the same university as the interviewees. We strived to avoid bias and maintain academic objectivity and criticism in our reporting and analysis by analyzing the data with multiple persons, by involving interviewees with different roles (i.e. policy makers, data managers, and researchers), by letting research participants review our results (including transcript reviews), by verifying the interview findings with multiple other data sources (e.g. policy documents), and by checking for alternative explanations of various findings. Furthermore, one of the authors presented the study’s findings in a keynote of an academic conference (ICTeSSH 2020) and collected feedback from peers. We recommend future research to conduct case studies at other faculties, universities, and research data infrastructure providers—thus case studies of other contexts. This would further increase knowledge of which institutional and infrastructural arrangements work best in which contexts. A valuable method for such arrangement development would be iterative prototyping and hands-on workshops with research data management policymakers, support staff, and researchers with both considerable and limited experience in research data sharing. This multiactor approach would take all relevant perspectives into account and further stimulate open research data sharing and reuse at an institutional and infrastructural level.

When researchers engage in serious and continuing research noncompliance and integrity violations, human and animal research subjects are put at risk and the trustworthiness of data and public trust in science are undermined. Serious research noncompliance is when researchers fail to follow federal, state, and institutional rules and regulations governing the conduct of research in a manner that increases risk to human subjects or compromises scientific integrity; continuing research noncompliance is when researchers fail to consistently and fully follow federal, state, and institutional rules and regulations related to the conduct of research (DuBois and Antes 2018; Dade, Olafson, and DiBella 2016; Shamoo and Resnik 2015). Research integrity violations include falsification or fabrication of data and plagiarism (Fanelli 2009; Office of Research Integrity 2020). In an ideal world, there would be no instances of research noncompliance or integrity violations. However, researchers are human beings who make mistakes, and these cases do happen. The research regulatory environment is complex, and executing protocols with the utmost precision and reliability requires project and personnel leadership and management skills that are sometimes lacking among researchers. The purpose of this study was to identify what institutional officials (IOs) perceive as common root causes of violations and the action plan activities they typically prescribe to researchers who have engaged in violations. Specifically, we focus on principal investigators (PIs), researchers who lead or direct research teams and who are the primary person responsible for the research being conducted. Accurately identifying and effectively addressing the complex factors that lead to research noncompliance and integrity violations is critical for reducing the risk of future violations. Incidents of research integrity violations and noncompliance are often multi-causal and interconnected. To illustrate, a lack of communication is often identified as a contributing factor to noncompliance. But upon closer examination, lack of communication can be caused by other more nuanced factors, including researchers being overextended, interpersonal issues among members of the research team, and a lack of training in effective communication strategies (DuBois et al. 2016; Neely et al. 2014). PIs are frequently held accountable for instances of research violations and noncompliance, even if they occurred in the absence of malice or by someone else on the research team they lead. Sometimes research noncompliance and behaviors that violate research integrity can be attributed to intentional actions taken by a researcher in a moment of poor judgment, such as when they are in a rush and opt to cut corners (DuBois et al. 2013). In such cases, institutions may choose to retain and remediate the researcher if they believe the risk of reoccurrence is low. In other more extreme and intentional cases, institutions may opt to terminate the researcher’s employment. Yet, in some extreme cases when the researcher is influential and well-funded, institutions may be more willing to retain and attempt to remediate the researcher. When incidents of research noncompliance and integrity violations occur, IOs must investigate and respond. The Office for Human Research Protections (OHRP), the Food and Drug Administration (FDA), and the Office of Laboratory Animal Welfare (OLAW) require research institutions to develop action plans to address researcher noncompliance or integrity violations (Department of Health and Human Services 2018; Kessler and Sullivan 1991; Office of Laboratory Animal Welfare 2018). It is often challenging to identify all contributing root causes and appropriate corresponding activities that mitigate the likelihood of reoccurrence. In the vast majority of cases, researchers will remain in their position and continue to conduct research (DuBois, Anderson, and Chibnall 2013). Therefore, it is essential to develop and implement a robust and effective plan to prevent future incidents. Plans, often referred to as a corrective and preventive action plans, or simply “action plans,” are commonly developed to achieve this aim. For an action plan to be effective in changing researchers’ behaviors and practices, several factors need to be taken into account. The first is that the root causes of the noncompliance or integrity violation are accurately identified and understood. Root causes can involve individual, group, and contextual factors (DuBois et al. 2016). Individual-level causes emerge from a researcher’s professional habits, attitude toward compliance, or other personal factors. Group-level causes emerge from research group management processes, interpersonal dynamics, and ineffective leadership practices. Contextual-level causes arise from the institutional environment and regulatory and funding systems (Mumford 2002; James, Burrage, and Smith 2003; Broome 2003; Davis, Riske-Morris, and Diaz 2007). The complexity and multileveled nature of root causes necessitates the development and implementation of a dynamic action plan. Action plan activities should be directly linked to the root causes identified to support researcher behavior change. Thus, it is important that IOs consider the full range of possible action plan activities that can be linked to root causes, especially those activities that are not on their “go-to” list of options. Our team’s prior work developing and leading the Professionalism and Integrity in Research Program (P.I. Program), a researcher remediation program that helps researchers identify and addresses root causes of violations, suggests there is variability in IO involvement in action plan development and in the content of action plans themselves (DuBois, Anderson, and Chibnall 2013; DuBois et al. 2018). This variability may be due to institutional resources provided to IOs, institutional norms, how frequently incidents of research integrity or noncompliance violations occur, knowledge and experience of IOs, and whether knowledge is transferred between IOs at the same institution. Understanding common root causes and corresponding action plan activities among IOs nationally can provide a fruitful starting point for evaluating the state of responses to compliance and integrity violations. We sought to answer the following research questions:

RQ 1: What do IOs perceive to be common root causes of research noncompliance and integrity violations?

RQ 2: What do IOs report as common action plan activities for remediating root causes of research noncompliance and integrity violations?

Materials and Methods

This study was approved by the Washington University Institutional Review Board (IRB ID# 202008019). The research team was led by social scientists with expertise in research ethics, responsible conduct of research, organizational psychology, qualitative research, lab leadership and management, and remediating root causes of research noncompliance and integrity violations (DuBois et al. 2015; Antes, Mart, and DuBois 2016; DuBois and Antes 2018; Antes, Kuykendall, and DuBois 2019a, b; McIntosh, Antes, and DuBois 2020; DuBois 2013; DuBois, Anderson, and Chibnall 2013; DuBois et al. 2018; DuBois et al. 2013; DuBois, Chibnall, and Gibbs 2015). The team’s collective experience conducting research on these topics provided a novel and robust foundation for informing the design of this study and interpretation and synthesis of findings. To answer our research questions, we conducted semi-structured, in-depth interviews with IOs who have experience addressing serious and continuing research noncompliance and integrity violations. Interviews were conducted with various types of IOs including: chief research officers (CRO), conflict of interest committee (COI) chairs, institutional animal care and use committee (IACUC) chairs and directors, institutional review board (IRB) chairs and directors, research compliance officers (CMPO), and research integrity officers (RIO). Interviews sought to understand what IOs view as common root causes and what activities are typically included in action plans at their institutions.

Recruitment and Procedure

A recruitment database was built using publicly available contact information, supplemented by snowball recruitment. Using a list of U.S. Clinical and Translational Science Awardees (CTSAs) and MD-granting medical schools, between August and October 2020, we searched these institutions’ websites for available contact information of IOs. IOs were recruited to participate in an interview via email invitation. Interested individuals clicked a link embedded in the email which directed them to a short pre-interview survey. Before answering survey questions, participants provided consent to participate. Survey responses were used to screen prospective candidates for inclusion in the interviews and gather demographic information and basic information on their experience with action plans. Participants were eligible to participate in an interview if they had a minimum of two years of experience handling noncompliance or integrity cases and participated in developing at least one action plan. After completing the demographic survey, a member of the research team scheduled interviews with participants. Participants completed a 30-minute interview with a member of the research team trained in conducting qualitative interviews. Participants were offered a $25 Amazon eGift Card for their participation. All interviews were conducted over Zoom, audio recorded, and professionally transcribed. Transcriptions were checked for accuracy and de-identified.

Materials

While all participants were asked the same general set of interview questions, slight modifications were made to the wording of the interview guide depending on the predominant research violation type that the IO investigated and adjudicated (e.g., human/animal subjects protection noncompliance or research misconduct). At the start of the interview, participants were asked about their experience handling cases of research noncompliance or cases of falsification, fabrication, and plagiarism (FFP). Then, participants were asked about common factors that contributed to cases, followed by questions about their action plan development procedures, content, and perceived effectiveness. Interviews focused on cases when IOs viewed the researcher as “redeemable” and were not focused on cases where employment termination was likely. Codebook Development and Data Analysis

Interview transcriptions were uploaded to Dedoose qualitative analysis software. Research team leads AA and TM led codebook development to ensure a rigorous coding process (Roller and Lavrakas 2015). A combination of inductive and deductive approaches were used to develop the codebook (Saldaña 2016a, b). Deductive coding involved structured codes based on participant responses to specific interview questions. Inductive coding involved reading through a subset of transcripts to identify views and attitudes about root causes and action plan activities that were spontaneously mentioned by IOs. Three study team members, AA, TM, and LR practiced applying the codebook to the same subset of interview transcripts and kept detailed notes during coding. They met to resolve coding discrepancies and revised the codebook as needed. The finalized codebook was applied to all 47 transcripts. Each coder was responsible for coding approximately one third of the transcripts and was assigned to review the code applications of another coder to ensure rater agreement across all transcripts. If there was a discrepancy, the coders met as a team to discuss and reach consensus on the appropriate code application. Once the coding was complete, the research team analyzed coded excerpts to identify themes and nuances within each code. Excerpt analysis involved reading through all excerpts for a code, summarizing the various IO perspectives within that code, identifying illustrative quotes, and meeting as a research team to discuss and reach a consensus about the overall meaning of the IOs’ comments captured within each code.

Results Participants

We interviewed 47 IOs who were chief research officers (n = 6), conflict of interest committee chairs and directors (n = 6), institutional animal care and use committee chairs and directors (n = 7), institutional review board chairs and directors (n = 13), research compliance officers (n = 6), and research integrity officers (n = 9). IOs often handled multiple types of noncompliance or integrity violations, including those that relate to: human subjects noncompliance (70%), research misconduct (62%), conflicts of interest violations (53%), animal subjects noncompliance (49%), and other types of noncompliance such as biosafety violations (19%). IOs had an average of 12.7 (SD = 7.43) years of experience handling cases of noncompliance or integrity violations and had participated in developing an average of 15.09 (SD = 18.65; median = 10) action plans. Additional demographic information is provided in Table 1. When asked about the extent to which action plans address institution and researcher needs on a scale of 1 (not at all) to 5 (a very great extent), participants perceived action plans to address the concerns of institutions (M = 3.94, SD = 0.73) more than the needs of researchers (M = 3.38, SD = 0.61).

Root Causes

Table 2 lists the root causes identified by IOs, sorted by the most common root causes mentioned overall. Generally, the most common root causes identified in interviews were similar across IO types with differences primarily in rank order. In what follows, we summarize IOs’ views about each root cause and supplement with illustrative quotes. Lack of Knowledge or Training—A majority of IOs indicated that noncompliance and violations of integrity arise in part from staff and PIs not fully knowing or understanding the rules and policies, or not knowing protocols well enough to implement them competently. They indicated that inadequate training is rather common, and the learning curve for understanding what supports the conduct of responsible and rigorous research can be steep. As indicated by these IOs: “…people just don’t remember the training and what they’re supposed to do and don’t ask…usually these are errors related to poor training or not understanding the training properly.” – CMPO.001 “Lack of training, lack of personnel, lack of trained personnel, and I guess not just trained but experienced. Not just ‘I know how to do it,’ but ‘I know how to do it and under different conditions’…How do you do this when everything’s perfect, and things are rarely perfect so you need people who are experienced with when things don’t go right [laughter] so that they can course correct.” – IRB.003 RIOs and CROs mentioned observing a lack of knowledge about how to cite sources properly, unawareness of what constitutes plagiarism, and misuse of statistics. Compliance officers and those officials who work with human and animal subjects researchers mentioned a lack of knowledge about procedures, such as how to properly document research, conduct certain aspects of a protocol, or manage data. Officials working in COI noted that PIs often do not know the policies and requirements for reporting conflicts. Another official noted that regulations change over time, and a PI might continue to follow old procedures and regulations. IOs also noted that sometimes PIs and staff do not realize that they must follow the specifications in a protocol exactly as written. As described by this IOs: “Not knowing you need to be explicit in the protocol about your plans, not just sort of following it…It was just a matter of explaining to them is it’s okay if you only want to do what you think you can do, but you should have written that in the protocol and said, ‘Here are the things that we may do, and we may not do them. It’s not important,’ because if you write in the protocol that this is what you’re going to do, then that’s what you have to do. It’s a matter of education, I guess, that people don’t or think that the protocol is something you have to absolutely follow.” – CMPO.002 Some IOs also noted a lack of researcher training on how to effectively run a lab, including how to lead and manage research personnel. One COI official stated: “…I would love to see the next generation targeted, so the new PIs that are coming in. Because I think part of the challenge is if nobody ever teaches you any different, you don’t know there is a difference. Part of it is bringing where people really start their career, and that point of starting their lab, what practices can they put into place. Because running a lab is running a small business. We don’t really train our PIs well on how to do that, how to deal with HR, how to deal with personalities, how to motivate people without berating them…I hope there’ll be more programmatic stuff on that end. Then continuous learning throughout.” – COI.006 Failure to Provide Supervision—IOs indicated that research team members and personnel often commit noncompliance and violations of integrity, not the PIs themselves. However, PIs are ultimately responsible for the conduct of their personnel. IOs suggested that a failure by lead researchers to provide supervision of the research and research team was a major contributing factor. Also noted was the need for staff to be able to obtain help and input from the PIs, but that sometimes staff might not seek this help or ask questions because they are under the impression that they are supposed to know and should not have to ask. As one IO noted: “…in one situation, there was inappropriately trained and supervised study coordinator. They didn’t understand how to follow the protocol and manage the data. They weren’t seeking help because they assumed they should understand it. If they were seeking help, that they might get in trouble ‘cause they didn’t understand it.” – IRB.013 Several IOs mentioned that they perceive researchers to assume that their staff know what to do and lab operations will work well without day-to-day supervision. Relatedly, one CRO mentioned a “laissez-faire” approach to management: “…It’s the lack of day-to-day oversight of individuals and their responsibilities…I think there is a lack of planning of the research teams that—in assuming that everything is gonna go well and everything…everyone’s gonna do what they’re supposed to do. I think there’s a certain amount of naivety with that and not assuming that there are some, perhaps, critical oversight functions that need to be put into place to ensure compliance. I think to a certain extent, some—a bit of the laissez-faire approach to management of, ‘Gee, I hope everything goes well.’” – CRO.001 When PIs do not actively and appropriately manage their research and staff, this creates the opportunity for mistakes and other research-related issues to remain unidentified and unaddressed. When these issues are not addressed in a timely manner, noncompliance is more likely to result. An example of a management issue noted by IOs is that researchers do not establish delineated research operations. Such operations include routinely checking on the implementation of protocols and specifying how to document information in lab notebooks, where and how to store and handle data, and maintaining version control. As one RIO stated: “…the thing that we see most often is when more inexperienced staff are kind of just left to their own devices and not being given appropriate training, or mentoring, or oversight, and they try to figure it out on their own. They do the best that they can, but then things kind of snowball and spiral, and they get out of control.” – RIO.006 Another management issue noted by IOs was failure to facilitate coordination among study team members and cultivate team-wide knowledge about the roles and responsibilities of everyone on the team. One RIO emphasized the need for researchers to be good project managers and capable facilitators of weekly meetings. RIOs and those involved with research subjects compliance noted that very large groups of students and staff without appropriate supervision can especially create problems for noncompliance. It was also noted that sometimes researchers’ priorities were focused more on data and productivity and less on lab member training and learning. Another supervisory concern that emerged was researchers not reviewing data and lab notebooks, and not questioning data: “…but if you are not a really good mentor and really looking at the data and the lab notebooks and questioning really ‘sexy’ results, then that contributes to not detecting the problem with work a grad student is doing.” – IACUC.002 Attitude or Personality—IOs discussed how certain attitudes contribute to noncompliance. Most prominent was PIs not viewing compliance as important, which led them to be inattentive to compliance-related matters. Failure to consider compliance as a priority resulted in behaviors such as forgetting to report conflicts of interest, not putting forth effort to ensure compliance with human or animal research protocols, and inappropriate responses when compliance officials brought errors to the attention of PIs. Closely related was an attitude that compliance officials do not understand the nuances of a PI’s research area and are unable to properly evaluate the conduct of the research. Similarly, some PIs view compliance officials as adversaries instead of being there to provide help. One IO talked about the importance of cultivating researcher buy-in to guard against negative PI attitudes or views about compliance: “They don’t really realize that compliance is truly an important thing and that we can get into a lot of trouble…if we aren’t compliant with the rules and regulations…by letting them know that we do need to follow the regulations…it gets buy in, from the investigators rather than we’re just not a place that signs the last page of the study, here, it’s approved. We actually take an interest in, and also make sure that we’re compliant and work with the investigators.” – CMPO.003 Overall, IOs noted that willful noncompliance or intentionally breaking the rules tends to be rare. IOs noted noncompliance by staff, not PIs, is relatively common. For example, staff may tell the PI they completed a required procedure when in actuality they did not. However, IOs did mention a contributing factor to noncompliance included investigators with difficult or strong personalities. They mentioned that some PIs can be dismissive, defensive, or arrogant and think they do not need to follow the rules: “Dealing with personalities like that and, ‘I’m a very important doctor. My research is very important. Therefore you should leave me alone,’ makes life a little more difficult sometimes.” – IACUC.002 Some IOs noted that arrogance and dismissing rules might occur alongside investigators prioritizing speed, efficiency, and productivity over everything else. Some RIOs referred to PIs “gaming the system” or believing they did not need to actually conduct the research because they were confident they knew what the results would be. One human subjects official mentioned PIs sometimes move forward with conducting research without approval because of frustrations about the pace of obtaining approval at their institution. Generally, IOs perceived that the majority of PIs want to behave correctly. As this IO stated: “…you have that opportunity to escalate and to teach, then I feel like that’s what you’re supposed to do. I tend to think that most people wanna…do the right thing, and wanna do a good job, and want their careers to continue.” – CMPO.005 Lack of Formalized or Clear Procedures—IOs noted that issues arise when teams do not have formalized plans and written procedures for executing studies. They lack clearly defined research operation practices or team procedures, such as how they are tracking information about human subjects protocols. Other concerning documentation practices IOs mentioned included: sloppy record keeping, improper documentation of research subjects procedures that were completed, failure to complete paperwork (e.g., participant signatures for informed consent), not using lab notebooks, and a lack of document version control. Another related issue raised by IOs, but especially for RIOs and CROs, had to do with improper data management (e.g., data storage and review) and that a lack of explicit procedures for properly managing data contributes to issues of data integrity. IOs also noted that a lack of a process for paper writing, especially when writing as a team, can result in accidental plagiarism or falsified figures. As several IOs illustrated in the following comments: “…something that I think routinely leads to reports of non-compliance. That is that the researchers, particularly maybe the novices, don’t put into place rigorous documentation management. Thus, that leads to problems. Typically, the consent forms are a good example. Using an outdated consent form. The question is, well, why—how did this happen? It’s because the research team didn’t put into place a rigorous way to manage their documents so that outdated ones get retired and the newest ones are immediately available to the research team. It’s just that structure and function that is oftentimes missing.” – CRO.003 “…we had some primates that were havin’ surgery, and the post-surgical care wasn’t bein’ well documented… the research technicians weren’t ensuring that the primate was actually taking—it was an oral painkiller, and the animal care staff had found the pills in the bottom of the cages on a few occasions. I just remember that was just continually a area of contention that the PI felt he was doing what he needed to do, and that sometimes those things happen. The veterinary staff was like, ‘No, this needs to improve, and this is inconsistent with the care expected of the guide.’” – IACUC.004 IOs often asserted that having written procedures and ensuring that they are followed is critical. How well the team coordinates the work is essential for putting into practice the required procedures within protocols. Explicit documentation of how protocols should be executed supports compliance with related rules and regulations and ensures data are of high quality. In addition, proper tracking and recordkeeping are essential. One IO noted the prevalence of data-related problems and how data management plans need to be established in all research teams: “The challenge with data management is it’s not a one size fits all solution. NIH is not entirely the right body in my opinion to fix [data management plans] because they’re a project-by-project basis. They’re not really getting at the core of how a lab is run from the beginning, and how we put processes in place to instill good data management practices with individuals. I would say almost 100 percent of the cases I’ve ever seen always involves some data problem, whether there’s a finding of misconduct or not.” –COI.006 Hurrying or Lack of Time—IOs indicated that sloppiness, shortcuts, or errors arise from researchers rushing to complete their research or perceiving a lack of time to do compliance tasks in the correct manner. As one IO explained: “Or a lot of times, it’s, ‘It was Friday afternoon, and it was 3:00. I had to get all this stuff done before 5:00. I went down. I was just in a hurry when I did this, so I didn’t push the cage in all the way’…It comes down to just being human and not always being 100 percent on their game. I would say that’s the vast majority of non-compliances that we have…” – IACUC.002 One RIO noted that time pressure related to visa timelines may contribute to noncompliance or integrity violations:

“Sometimes, they’re here on a short-term visa. Not only do they have the inherent time pressures that we hear from all of our young researchers…they have this added pressure. They are truly only here for a set period of time, and they must produce because they only have this opportunity.” – RIO.007 Overextension—IOs described the perils of researchers having too many responsibilities and projects at once. They indicated that overextension can lead researchers to be inattentive and even confuse research protocols. As one IO illustrated: “…sometimes PIs get overwhelmed with having a large number of protocols…I’ve sometimes seen PIs who sometimes forget which procedure is approved on which protocol…I’ve seen this happen on more than one occasion.” – IACUC.001 They noted that overextension causes researchers to neglect critical activities, such as training their personnel and managing their projects. In the words of one IO: “Many of them get overextended. You have too many projects…You get PIs, they have three or four or five or six or eight R01 grants. I mean, my gosh. No lab management.” – RIO.004 Another IO expressed a similar sentiment: “I think it’s not paying attention and conflict of commitment, being over-committed, not overseeing—under most of the cases is not overseeing the research that’s being conducted by staff and just not paying attention to detail and being sloppy.” – CRO.002 With regard to conflicts of interest, an IO noted that some investigators get so many outside invitations that they fail to keep track of all external sources of money. This results in reporting errors or omissions: “Some people really get so much money outside of their job that they can’t keep it straight…I’ve seen it a lot again with our very prolific investigators who are world renown. They get sought after by a lot of companies to talk or give opinions, advice, and so, I’ve been told before, ‘I just forgot that. I missed that one. I forgot about that $5,000 from that company.’” – COI.001 Academic Pressure—IOs described academic pressure as a contributing factor to research integrity and noncompliance violations. They indicated that the pressure on researchers to obtain funding in a competitive environment, pressure to publish, and relatedly, pressure to earn tenure are salient factors that shape researcher motivations. In the words of one IO:

“…I think there’s a lot of incentive problems in academic research that can cause some difficulties of cultures within laboratories…the so-called publish or perish, but the idea that only the sexy science gets published, which motivate[s] people incorrectly…negative data doesn’t get published…hypothesis bias becomes more of an issue. It can erode the strong practices we have in labs. There’re frankly is just too many Ph.D.’s. The competition is high within laboratories to produce.” – COI.006 Another IO felt similarly, highlighting the added pressures of soft dollars research positions: “I think a lot of it has to do with people who are on soft money are desperate. You have to get that next grant. You have to get that next paper because, if you don’t get that next paper, you don’t get that next grant. If you don’t get that next grant—it becomes a cycle.” – RIO.008 Some IOs noted that pressure, especially in a researchers’ early career, causes significant stress and can lead to bad decisions and misbehavior: “I think pressure to get grants. That would be number one. Number two would be pressures around promotion…Number three would be the personal pressure that individuals put on themselves to perform at a certain level…if you’re not getting grants, don’t even think about tenure promotion…the not getting publications are gonna essentially push you off into a corner. Folks will bend the rules. In my experience, it starts out with little things, and then it gets to be bigger things.” – CRO.004 One IO mentioned the need for providing support and tools for researchers to better handle performance pressures: “People I think are generally well intentioned. There’s so much pressure from so many different angles—from the institution, from the funding agencies, from the public, from mentees, from colleagues. It can be very challenging for our investigators to stay on top of everything. To the extent that we can develop tools to support them in their efforts and to help them understand…how can we take some of the burden off of them and help them do what they’re supposed to do?” – IACUC.007 Communication Problems—IOs noted that research team members are often the ones who commit research noncompliance or integrity violations—not the PI. Communication breakdowns are a key contributor. Miscommunications included researchers not clearly communicating with staff or not communicating at all. As this IO noted: “Well, it’s mostly calling their attention to they’re not as smart as they think they are, in terms of being able to deal with the regulations, ‘cause it usually is a matter of the regulations…it’s usually calling their attention to the fact that their communication was a little ambiguous…they kinda see it—that they need to think in terms of how the researcher or how their students think, rather than how they think.” – IRB.012 There were also instances where staff did not communicate well among themselves. Miscommunication between research team members and compliance groups was also an issue. As one IO commented: “Honestly, it’s just communication…a few mouse cages were missing food and water. It was a lack of communication within DLAM [Division of Laboratory Animal Medicine] staff. These people are all really dedicated and good. There’s no way they wanna do that. A thought B was taking care of them, and B thought A was on top of that. Again, lack of communication.” – IACUC.003 One RIOs raised concerns about labs in which teams do not communicate openly about data: “I mentioned toxic lab environments—right?—where people don’t freely talk about and review data.” – RIO.005 Institutional Environment and Systems—Several IOs mentioned that sometimes the institution did not provide adequate guidance, resources, support, and tools for researchers. They viewed this factor as contributing to noncompliance. IOs mentioned lack of adequate education, mentoring, and monitoring for researchers as contributing factors. As one IO articulated: “…just a lack of appropriate infrastructure. Whether it be education for the research community about what’s expected or lack of modern tools or lack of resources available to respond to issues that do arise. That’s probably one of the more common things that I’ve seen.” – CMPO.006 Promotion criteria, the demands put on faculty, and the need for more balance in their lives also emerged as prominent factors: “Part of it is thinking about…rethinking promotion criteria, rethinking what the requirements are of faculty, how we can support faculty to have more balance in their life, and to run their lab productively, but not be a machine.” – COI.006 Some IOs noted that cultures within departments at institutions contribute to research integrity and noncompliance violations: “…looking at it more in an institutional context and determining…is there something systemic in the culture of the department [that] led to this.” – RIO.002 One RIO specified instances where institutions permit bad actors and fail to take swift action against bullying behaviors by researchers. To address this concern, leaders within the institution should simultaneously address problematic behavior while promoting a culture of compliance and integrity throughout the institution. As they noted: “…climate and culture within individual labs and institutions permitting bad actors…If you’re a bully running a huge lab…you have some poor kiddo who is inclined to just take a shortcut to get to the results you need because you’re breathing down their neck all the time and asking for a certain result—I think that poor mentorship and poor institutional response or lack of institutional responsibility for addressing those people because they bring in a lotta money is definitely a contributing factor.” – RIO.007 Desire to Produce a Particular Result—Some IOs mentioned researchers’ desire to obtain good results can lead them to not build in safeguards and pay as much attention as they should. IOs also mentioned that researchers can sometimes put pressure on their research team to get specific results. For example, one IO said: “Then the faculty, in turn, put that pressure on their lab—their research group to get those results. I think that external pressure can be a big factor.” – CRO.006 Also mentioned was the notion that researchers’ desire for good results can lead them to not questioning results from lab members that look too good to be true. Similarly, researchers can be so confident that results are correct, they fail to review the data. In the words of one IO: “…sometimes, it is they’re just so sure they’re right that either they’ve convinced themselves that they’re so right, or they’ve convinced themselves that there couldn’t possibly be another outcome. Therefore, why waste the time?” – RIO.002 Lab Culture—IOs, and especially RIOs, mentioned how toxic lab environments contribute to misconduct. Toxic lab environments are characterized by research team members not openly sharing or talking about data, hyper-competitiveness, and bullying. As this IO describes: “…if it involves the entire lab culture, which usually starts with the PI, those are the more difficult ones…Was it some issue in the lab with people not getting along…you can’t turn a blind eye to some issues that are just all around. For those other issues, let’s say sexual harassment or other types of harassment, bullying, we will refer to other offices for their own internal investigations.” – RIO.003 IOs also noted that researchers create bad research environments when they do not model responsible behavior:

“It was all Photoshopped, but he was making a lab that was known that that was encouraged. He should be even more responsible for encouraging a next generation to do this…he should be held to a higher standard for telling others to do something than a young student who…doesn’t necessarily know better.” – RIO.008 Cultural Background—IOs described cultural background, including differences in assumptions and norms related to the conduct of research, as a contributing factor to noncompliance and integrity violations. They noted that some international researchers experience language barriers resulting in miscommunications with their teams, supervisors, or with compliance groups. Language barriers also create additional pressures on researchers, which might, in turn, contribute to misconduct. As one IO pointed out: “…not always is English the primary language in some people in labs. Sometimes there’s a communication barrier there to where we realize he says one thing, and this person interprets it another way. That’s a challenge.” – IACUC.003 Researchers from non-dominant cultural backgrounds may have different norms and practices for conducting research. IOs mentioned that issues related to animal care, plagiarism, and workplace behavior may differ as a result of cultural background. Due to differences in training or upbringing, researchers who are not from the U.S. may not understand research norms and expectations in the U.S. Another IO explained: “I’ve dealt with individuals that have plagiarized…I’m not sure they’ve always, necessarily, understood or appreciated—never learned or were never trained properly in how you cite data, how you cite other people’s manuscripts, how you give credit to other authors, that type of thing.” – CRO.004 Personal Life Factors—Some IOs identified factors in a researcher’s personal life as contributing to research noncompliance or integrity violations. This was especially true for personal issues that cause pressure, stress, and distraction. Financial issues, divorce, mental health issues, and substance use were mentioned as potential personal factors. One IO stated: “Was this somebody that had another issue in their life that was leading to a significant lapse in judgment? Maybe they’re going through a very messy divorce. Maybe they have an alcohol or drug problem.” – CRO.004 Lack of Staff—Another root cause that some IOs identified was researchers whose teams lacked adequate staff support. When research teams are understaffed, there are not enough people to assist with activities that need to be done. Some IOs mentioned that not having a lab manager to help provide oversight contributed to cases. As this IO illustrated: “I think what makes it difficult…is the amount of time they actually have to spend with their labs. If they don’t have a really good, seasoned lab manager that’s doing the on-the-ground, day-to-day management, that is what cripples them from getting people to do what they need to do and follow things.” – IACUC.002 Problematic or Misunderstood Protocols—IOs mentioned protocols themselves are sometimes problematic because there are too many to oversee and researchers get them confused. Or, sometimes the protocol might be too complicated or too strict, such that it is difficult to follow. IOs noted the need to write protocols that better accommodate the study team’s needs and plans. “…sometimes PIs get overwhelmed with having a large number of protocols…I’ve sometimes seen PIs who sometimes forget which procedure is approved on which protocol…They get overburdened, and they’re just really not—again, they need somebody to help manage the situation.” – IACUC.001 Finally, one human subjects official mentioned the problem of researchers not fully understanding the protocol, especially differences between following the research protocol versus standard of care: “Sometimes, it’s a matter of confusion between standard of care practice and following the research because what you’re doing for standard of care is different. This was the investigator trying to apply standard of care to a research protocol that had a more stringent procedure in place for a procedure, where in standard of care, you had some more leniency on the way in which you addressed a number of attempts…” – IRB.013

Action Plan Activities Table 3 lists the action plan activities that IOs identified, sorted by the overall most common action plan activities. While many action plan activities were commonly mentioned by multiple types of IOs, there is some variability in action plan activities among IO types. For example, the most common action plan activities among RIOs were requiring oversight or mentoring of the researcher and providing researchers with training to develop new professional skills. The most common action plan activities among IRB and IACUC officials was retraining in compliance or research integrity and having IOs follow-up and provide hands-on involvement. We summarize what IOs thought about the different action plan activities and provide illustrative quotes. Retrain in Compliance or Research Integrity—When asked about specific action plan activities, the most common activity mentioned by IOs was requiring researchers to be retrained in compliance and research integrity. PIs, and in some cases their personnel, need to learn or relearn research policies and expectations for upholding compliant, rigorous, and responsible research. As one IO noted:

“…we definitely have the education component involved. We require more training, we sometimes say, okay, we know everyone has to be trained, but we want them to do the training again even though it doesn’t expire for another year, or we only require core modules, but there might be some additional ones.” – IRB.010 Sometimes, researchers may be required to sign off that they understood the training and paid attention; this is done in an attempt to cultivate accountability and encourage them take the training more seriously. As indicated by this IO: “…investigators either forget the regulations or just don’t follow the regulations because they don’t know them know well enough…unfortunately, when training comes up people take it lightly. They don’t pay attention to the training…People turn on a training, just let it run in the background, take the test and pass it…and just don’t pay attention to the regulation. A lot of times we’ll institute a repeat of that training… We make sure that they have a sign off at the end, where they’re specifying, yes, I understood this, and I’m going to follow it.” – CMPO.003 Hands-on Involvement and Follow-up—IOs noted that some PIs and their team members need additional personalized instruction from compliance support persons to understand how to do procedures correctly and compliantly. Additionally, IOs sometimes engaged in conversations with PIs to ensure they understood the seriousness of the issue, or to help them identify the underlying issues in their labs and approaches to addressing them. IOs also discussed following up or checking in with researchers after a period of time to ensure that action plan activities are on track and that activities are being completed appropriately and sufficiently. Doing so cultivates accountability on the part of the researcher and gives the IO reassurance that the action plan is having the intended effect on researchers’ behavior and that correct research procedures are in place. Check-ins also provide opportunities for PIs and lab members to ask questions and serve as a reminder of the importance of compliance. Follow-up may occur once, multiple times, or routinely (e.g., by embedding a regulatory specialist within the study team to help them prospectively on a daily basis). In their involvement, IOs noted that they typically avoid taking a punitive approach and instead, adopt a more support-oriented approach in their communication and interactions with researchers and their teams. However, in certain cases involving very serious noncompliance or conduct, a punitive approach involving additional university leaders (e.g., deans, department chairs) may be appropriate. As these IOs stated: “You have somebody go through it with ‘em…and show ‘em exactly what it is, do it again, and oversee them. I think it depends on how egregious it is whether or not it’s just, do you take more of a…passive role…Passive in the sense of okay, you leave it on them to go and take the training again, or you actually put someone over them in the lab, so that they can watch what’s going on, or in the animal care facility to watch what’s going on and to monitor.” – CMPO.005

“We might go back every three months for a while. You can see, you could observe, them addressing the problems and fixing ‘em…Are they getting it right? What do these forms look like now? Are they handling this now? Are they getting consent the proper way? We’ve got a good chance for a follow-up…You have to go back and check them.” – IRB.006 Require Oversight or Mentoring—Researchers can be required to receive formal mentoring or oversight by a peer, department chair, or other institutional official. Oversight can take the form of mandating researchers to submit all research proposals and publications to be internally reviewed prior to submission. Mentoring can take the form of partnering with another person who can act as a “second set of eyes.” Duration of mentorship and oversight can vary widely depending on the incident, recurrences, and researcher needs. As these IOs illustrated: “…a lot of these situations come from lack of oversight, so when they come to our attention, we might assign additional oversight moving forward. Maybe that’s more meetings with their chair or someone to oversee what’s really going on in the lab. Someone that’s not conflicted or related to the research might look at data to see what’s going on on the ground…” – COI.003 “…this person needs a mentor in order to ensure that the issues that were uncovered during the course of the misconduct proceeding are addressed, not just in this specific situation, but that there is infinite mentorship available to this individual until they become a tenured professor for them to be able to touch base to ensure that their decision-making is appropriate in certain circumstances.” – RIO.007 Halt Study Protocol—IOs described temporarily or permanently stopping data collection and research activities on the protocol(s) of the researcher in question. This can occur during the investigation process or as a result of investigation findings. Restarting research activities after a protocol is halted might only occur after the researcher accepts personal responsibility and after action plans are developed, initiated, and underway. Sometimes this option is only for extenuating circumstances (e.g., when research subjects are at risk) or a last resort. For repeat offenders who do not correct their noncompliance, their ability to conduct research is suspended long-term. One IO noted: “Of course, we can actually withdraw their ability to do some kinds of experiments for a certain time period if it gets bad enough, and we’ve done that. Typically before we do that, we tell the person, ‘Okay, we’ve accepted your corrective action plan this time, but if this happens again you’re not gonna be doing this research project anymore.’ That usually gets people’s attention, too.” – CMPO.001 Another IO expressed a similar sentiment:

“We’ve had emergency meetings to discuss an issue where we’re like, ‘We just have to stop this study, put the brakes on, and fix things now.’…it’s nice just to have the full IACUC behind these decisions…I am all for stopping a study if something is seriously wrong, like the animals are definitely in danger.” – IACUC.006 Improve Staff Training—IOs indicated that often it is not only the PIs who need to receive training to address noncompliance, but staff conducting the research do as well. This action plan item includes providing new or additional training for research personnel to ensure they receive proper information and guidance and develop the necessary skills and knowledge to properly conduct rigorous research and follow regulations and study protocols. Staff training is especially effective in cases of unintentional misconduct or noncompliance due to gaps in knowledge or skills. Training can cover a range of topics including specific regulations or best practices for data management. Training can also provide staff with knowledge about where to find answers to questions. IOs indicated that sometimes they or compliance support personnel may provide this training, or the PI may implement new training. For example, one IO said: “…the PI may say, ‘Okay, we’re conducting training with our staff to understand the protocol more.’ Okay. If you’ve conducted training, do you have a log of the signatures, or how did you track to make sure that…they actually completed the training…the last check box is always, all right what’s the prevention, what are you doing to prevent? Like I said, it’s almost always training and education. We always say, fine if you’re giving [research staff] more focused training, just let us know what are you training them on and give us a documentation that they completed…” – IRB.002 Another IO stated: “…they need to have a conversation with their research team about the oversight and make sure that the research team understands that there was either an inadequate report of interests or that the management…The conflicted investigator is asked to discuss that with their research team, and they can pull [institutional officials] in if they would like that help.” – COI.004 Terminate or Involve Human Resources (HR)—Termination was reserved for serious cases, especially when researchers adopted a combative attitude, omitted information, lied, or if the noncompliance or misconduct was intentional. Several IOs mentioned involving HR in disciplinary actions, especially to document behavior in the researcher’s HR file. In cases when termination was pursued, the institution’s legal team was often closely involved and the researcher had the right to appeal the decision. Some IOs mentioned that to their knowledge many institutions do not inform researchers’ subsequent employers about sanctions, terminations, and HR disciplinary matters, and other IOs cited concerns about this practice and advocated for sharing information regarding problematic researchers.

“I always go and talk to the faculty member…I point out that, “If you’re up front, truthful at the very beginning of the process, if it turns out in the end that there was an untoward action on your part or recurring issue, the likelihood of separation from the university is much lower, and we can develop a plan. We can move forward. This is not necessarily career-ending…If you fail to be forthright and fully help the process and help the committee…it will make it much more difficult for the university to work with you at the end of the process should it not go in the direction that you would like it to go…”” – CRO.004 Require Retractions or Corrections—Action plans sometimes require corrections to be made to the scientific record. This can include retracting publications and conference papers, and correcting COI disclosure statements in the scientific record. IOs mentioned that action plans should specify the timeline and exactly how the researcher needs to work with the scientific entity (e.g., journal, professional society, funding agencies). IOs also requested that researchers provide documentation of completion. Issuing retractions or corrections was most common among RIOs and COI officials. In an effort to prevent future retractions and corrections, some IOs mentioned that they require copies of presentations, publications, or disclosure statements to be shared with them prior to the work being sent for review. “We make them provide us with copies of presentations and publications so that we can not only review…content, but we can also make sure that disclosures are in there. For publications, they do need to disclose it to us before they are…sent to a journal. If something has already been published and we are finding that there’s a conflict later, we make them notify journals.” – CMPO.004 Develop New Professional Skills—IOs described requiring PIs to develop new professional skills through training. The primary intention was to improve the research management knowledge and skills of the researcher who engaged in noncompliance or integrity violations or supervised team members who did. Professional skills development often occurs in formal training programs (e.g., P.I. Program) (DuBois et al. 2008; DuBois et al. 2018), but sometimes occurs informally (e.g., pairing with an exemplar to learn from them). These IOs note the importance of developing research team management skills beyond formal scientific training: “…determine when is the last time this respondent…took some professional training in…research management, project management…I’m always looking at project management skills…Are they trying to maintain some level of current skillset in project management?…I think those are critical skills that all researchers need to be current with. If there’s none of that going on and our investigation committee is looking at what I call an environmental scan or an environmental assessment of how the lab, for lack of a better term, is being managed on a daily basis and determine whether or not there are some common-sense elements like continuing education in areas that may not traditionally be thought of as opposed to those that are very technical.”

“…there was also education to the investigator and mentorship provided to the investigator on how to manage clinical research stuff and ensure compliance by their research staff. Make it clear that the work of a study coordinator was still their responsibility as a PI.” – IRB.013 Require PI to Draft Action Plan—Oftentimes having the researcher participate in the development of their action plan, including generating solutions for addressing root causes of the issue, helps cultivate ownership of and engagement with the corrective action plan. Sometimes action plans are developed collaboratively between the researcher and IO, with IOs providing requests for modifications to the plan as appropriate (e.g., to ensure that the researcher incorporates solutions that will effectively address specific root causes). Other times, IOs provided a list of action plan options for researchers to choose from, which may not be as effective as a more collaborative approach to determining action plan activities. As noted by this IO: “…we tend to get the investigator and the study teams engaged in the process of developing their corrective action plans. With that engagement comes an increased awareness of the potential problem areas, and a greater sort of understanding of what steps need to be taken to prevent violations and serious noncompliance from occurring…the corrective action plans tend to be quite successful.” – IRB.008 Create New SOPs or Procedures—Some action plans require researchers to develop written standard operating procedures (SOPs) or checklists for systematically organizing and describing lab operations and policies (e.g., for data management, record keeping, human and animal subjects procedures). SOPs support less individual variation among the research team and more standardization in the conduct of research. Asking researchers to develop SOPs also helps them identify problems, factors that led to these problems, and a plan to prevent the same mistakes from happening in the future. While IOs may recommend best practices to researchers as the SOP is developed, it is critical that the researcher takes an active role in developing SOPs given that the SOPs will be tailored to the individual research project and team. Some IOs follow up with the researcher to make sure the SOP is implemented and working as it should. “We had that PI write down policies and procedures for everything they do in their research program. They had to have a policy on how they do consents, a policy on how they store samples, everything. Then how you would not start a study without creating a regulatory binder and enrollment log and just everything. You have to have the paperwork for everything. They had to submit all their SOPs to us to review. Then we allowed them to resume research.” – CMPO.002

Hire or Appoint a Lab Manager or Add Staff—As part of the action plan, PIs may be required to hire someone, or appoint and train someone already within their research team, to help manage lab operations and compliance. Doing so helps improve oversight of research and can help researchers cope with being overextended. Lab managers should ideally be experienced, be able to quickly spot areas where issues might arise, and quickly bring issues to the attention of the researcher. Beyond hiring a lab manager, researchers should consider hiring additional staff who can, in addition to their regular roles, serve as a backup lab manager. In cases where hiring a full-time lab manager is not feasible (e.g., lack of funding or institutional support), researchers can consider hiring a part-time person and share the cost with another researcher. “…we required that the investigator not conduct any research until he had a clinical research coordinator to support him. He was trying to do it all on his own and thought he could, and he just—he couldn’t. He needed the support…find more systematic approaches to help them organize, ‘cause usually it’s a matter of disorganization or improper oversight are usually the two things that we see, improper data management and improper oversight.” – RIO.006 Suspend Ability to Mentor—Researchers who fail to oversee and mentor lab members properly sometimes have their ability to mentor trainees suspended. In certain cases, the ability to mentor is temporarily withdrawn, and other times permanently withdrawn. IOs often reinstate a researcher’s ability to mentor once researchers take steps to improve their mentoring skills (e.g., training lab members well, reviewing work of lab members), such as mentor training. “…there has been some question as to whether or not they had the required expertise to properly mentor graduate students. We have had situations where we have temporarily suspended their ability to mentor graduate students until they received a certain level of training. Then even once they are allowed to mentor the graduate students, then we had their department chair overseeing that mentorship for a period of time to make sure that the researcher got back on track.” – RIO.002 Consolidate or Limit Protocols—Some action plans temporarily limit the number of active studies a researcher can have ongoing at once, especially in cases where their noncompliance affected a large number of active studies. It is sometimes helpful for researchers to pause and figure out how to fix noncompliance prior to resuming research at full speed. “…limiting the number of studies they do for a while because there’s a whole series of things depending upon the action that occurred… One individual…was suspended from overseeing animal studies for quite a while.” – CRO.002

Other Action Plan Activities—Beyond the common action plan activities identified by IOs, there were unique activities or approaches that were not mentioned frequently enough to be included in the codebook and were mentioned by only one or two IOs. In cases of conflicts of interest, IOs recommended two actions: 1) in cases where a researcher fails to report consulting income or goes over the institution’s limit, require the researcher to pay back money given by industry and/or ban them from receiving industry money for a particular amount of time, and 2) require the researcher to disclose and have a conversation with their research personnel about their conflict of interest, which is intended to empower lab members to speak up if the researcher is doing something wrong. One IO mentioned the possibility of requiring the researcher to explain to a faculty committee why they violated the rules and why they should be allowed to continue conducting research. This approach may be best used in extreme cases where researcher attitudes are challenging and when public embarrassment of a violation is likely to result in a shift in researcher outlook. Another approach recommended by IOs was to have an institutional leader (e.g., department chair, dean) send a letter of reprimand to the researcher that conveys the importance of working with IOs. In cases where research was conducted without IRB or IACUC approval, IOs required researchers to delete or discard their data, especially if it contained protected health information. One IO mentioned the possibility of having noncompliant researchers serve on the IACUC or IRB, but they did not necessarily endorse this approach.

Discussion

Institutional officials are charged with the challenging task of addressing research noncompliance and integrity violations when they occur. Researcher remediation can be time-consuming and expensive and requires knowledgeable and patient IOs who have a desire to rehabilitate rather than punish. Ideally, identifying root causes of researcher noncompliance and integrity violations acts as a basis for formulating effective action plans that help prevent future occurrences of these cases. Root causes reflect a complex interplay of individual, group, and contextual factors. When action plan activities align with the multifaceted and often interconnected root causes identified, they are more likely to be effective in changing researcher behavior. We sought to understand the experiences of IOs who work at research-intensive institutions to consolidate IO knowledge about root causes they have identified and activities they have included in action plans.

Disconnect between Root Causes and Action Plans

Findings from IO interviews suggest that many of the most common activities included in action plans do not directly address the root causes identified by IOs, with the exception of a lack of researcher knowledge or training. Moreover, it is worth noting that many of the top root causes identified by IOs may be the result of less commonly identified causes, such as workplace culture and overextension. Our team’s prior work and the work of other scholars on researcher wrongdoing and remediation and factors that support the conduct of rigorous research suggest that a lot of what is known to be effective in cultivating researcher behavior change is not being effectively applied in institutions’ approaches to remediation (DuBois et al. 2016; DuBois et al. 2018; Antes et al. 2016; DuBois 2014; DuBois et al. 2015; Antes, Mart, and DuBois 2016; McIntosh, Sanders, and Antes 2020; Neely et al. 2014; DuBois, Chibnall, and Gibbs 2015; Gibbs 2013; Bandura, Barbaranelli, and Caprara 1996; Bandura 1999; Moore and Loewenstein 2004; Irwin 2009; Association of American Medical Colleges 2008). That is, action plans that simply halt study protocols or require oversight of researchers are unlikely to fully address disengagement from compliance, overextension, or a lack of appropriate procedures for conducting research and managing data. This incomplete alignment of root causes and action plan activities helps explain findings from the pre-interview survey indicating that IOs perceive action plans to better address the concerns of institutions compared to the needs of researchers. Furthermore, the disconnect between certain root causes and action plan activities points to the need for IOs to reexamine assumptions about the efficacy of current approaches to action plan development and implementation. Most IOs we interviewed had the attitude that most PIs want to do the right thing. Many IOs also noted that action plans should be approached as a bidirectional partnership with the PIs so that PIs can learn and be supported as they take steps toward addressing the problems in their labs. This partnership mentality is likely to prove important to fostering PI growth and change. Of course, PIs must also meet IOs halfway and engage with them as partners.

Rethinking Researcher

Training Training researchers to develop new professional skills was mentioned by just over a third of IOs. Yet, not all researcher training is of the same caliber, and certain training programs may not be well-suited for developing particular skills or changing particular behaviors. That is, some trainings, especially generic, informal, or off-the-shelf trainings, can be limited in the extent that they cultivate lasting behavior and attitude change. For any training to be effective, IOs must identify a framework to fully understand what researcher cognition and behaviors they are trying to change and what mechanisms are best suited for cultivating this change. For example, training focused on professional decision-making, priority-setting, and leadership and management skills may be especially important for addressing self-serving biases and cognitive distortions of researchers and a lack of effective researcher supervision in the lab (DuBois et al. 2013; DuBois, Chibnall, and Gibbs 2015; Antes, Mart, and DuBois 2016). Professional coaching can help researchers to re-examine priorities and challenge thought patterns that undermine compliance. Decision-making training can assist in the development of skills such as emotion and bias management, anticipating consequences of actions, and seeking help from appropriate sources. Examples of effective leadership and management practices that foster effective communication and coordination in the lab to support compliance and research integrity include building relationships with lab members, encouraging lab member engagement, addressing conflict, holding regular and effective team meetings, reviewing the work and data of lab members, and addressing mistakes constructively (McIntosh, Sanders, and Antes 2020; Antes, Kuykendall, and DuBois 2019b; Nembhard and Edmondson 2006; Rogelberg 2018; Antes, Mart, and DuBois 2016; Antes,

Kuykendall, and DuBois 2019a). While these types of researcher training may not be immediately apparent options that IOs consider, the skills developed during these trainings can help researchers restructure their approach to leading research in a way that can help them address more obscure root causes that may not be as apparently actionable, such as a lack of time and feeling overextended. Moreover, further research is needed on the role that lab culture and interpersonal dynamics within research teams play in research compliance and the trustworthiness of research. For the majority of IOs we interviewed, retraining researchers in compliance or research integrity appeared to be a common approach for addressing various root causes. If prior knowledge-focused researcher training had failed to result in application of that knowledge, it is important for IOs to identify why. Is the educational format (e.g., online module) not effective? Is the knowledge not retained because it is not viewed by researchers as important? Is the knowledge so complex and challenging that researchers need access to knowledgeable experts as questions arise? Does training need to focus more on educating researchers about whom to contact at their institution with questions and where to find information on demand? By asking these questions, IOs can begin to rethink and revamp training that is included in action plans. Another related approach identified was improving staff training. While providing knowledge-focused training to researchers can often be an important part of the equation, IOs should remain cognizant about more proximal root causes and associated action plan activities that supplement retraining. For example, lack of knowledge might have contributed to the problem, so retraining in basic knowledge may be appropriate. However, a deeper cause might be failure of staff to consistently apply knowledge from training when performing compliance tasks. Therefore, actions to address this cause might include staff training on specific protocols and subsequent development and implementation of a lab SOP, compliance checklist, or job aid. Establishing formalized SOPs and providing quality and routine training to lab members can help address matters related to failures in providing supervision. Staff training can also help support consistent use of SOPs and greater standardization in the conduct of procedures, which support research rigor and compliance. Training can vary in terms of how time-intensive it is, who delivers it, how customized or standardized it is, and how formal or informal it is. Due to limited resources and time, it may be tempting for IOs to only require passive trainings (e.g., online CITI) that are less likely to yield behavior or attitude change compared to trainings that involve active learning and engagement with training content. More extensive, involved trainings that require active engagement by trainees may better support changes in their behavior. Trainings that are tailored to identified root causes with specified learning objectives and measurable, documentable outcomes are recommended.

Better Preparing Institutional Officials IOs across institutions have diverse levels of experience and education in fields ranging from the natural sciences to law. While some IOs have handled several cases of research noncompliance or integrity violations, others have not had this level of exposure to such cases. As such, it should not be assumed that all IOs have the knowledge, skills, or resources to deal with complex cases of research noncompliance or misconduct. IOs would benefit from training on the multifaceted root causes of failures in research compliance and integrity, the development of effective action plans, and principles of behavior change and implementation. We encourage IOs to think about action plan activities that cultivate training transfer and consistent performance of essential practices, as these activities must facilitate researchers’ application of knowledge and skills to their actual work (Broad and Newstrom 1992; Burke and Hutchins 2007; Grossman and Salas 2011). For example, follow-up and hands-on involvement of IOs may help cultivate accountability, and can empower researchers to ask questions and work through problems when integrating new knowledge into their day-to-day lab operations. To support action plan success and the effectiveness of IO involvement, quality collaborative relationships and healthy rapport between IOs and researchers are needed. Adversarial relationships may spur resentment among researchers and make them less likely to adhere to action plans and feedback from IOs. Cultivating researcher buy-in can support successful action plans; researchers need to understand the rationale behind action plans and the importance of their implementation. To illustrate, mandating compliance-focused training alone is unlikely to cultivate an intrinsic desire in researchers to prioritize compliance and cultivate accountability and long-term behavior change. Framing action plans as supporting career development and success, instead of exclusively as a disciplinary tool, can make it more likely that researchers view the action plan process favorably and become intrinsically motivated to implement it. Therefore, IOs may need training on how to best communicate and cultivate strategically positive relationships with researchers at their institutions in general and in cases involving researcher remediation. Different IOs may have different perspectives on issues related to addressing noncompliance and integrity violations. While the most common root causes were mentioned by all IO types, there were certain root causes not mentioned whatsoever by certain types of IOs. For example, IACUC officials never mentioned the root causes of institutional environment and systems, desire to produce or overconfidence in a particular result, and lab culture. COI officials never mentioned the root causes of cultural background, personal life factors, or lack of staff. The differences in root cause identification among IOs suggests there may be variability in the level of understanding of root causes by IOs, or there may be some risk factors that may be more common in some domains. A similar pattern emerged with action plan activities. To illustrate, CROs were the only type of IO to mention consolidating or limiting protocols as an action plan activity. This suggests some IOs may consider a greater breadth of action plan activities than others, or certain action plan activities may lend themselves better to addressing certain types of noncompliance and integrity violations than other activities. Alternately, IOs may have their “go to” root causes and action plan activities that they reflexively identify and recommend. Receiving training on root cause identification and action plan development can help IOs be more intentional in thinking about alternate root causes and action plan activities that they would not have otherwise considered.

Limitations

Our findings provide valuable insights but are not without limitations. We only interviewed IOs from CTSA and MD-granting institutions, which are highly research intensive. Institutional size, resources, and culture may affect the nature of root causes of noncompliance or integrity violations and the action plan activities prescribed. Findings may not generalize to smaller and less research-intensive institutions. Another limitation worth noting is that we only report on root causes and action plan activities explicitly identified by IOs, which may not reflect a comprehensive scope of possible root causes and action plan activities. Moreover, some IO responses may reflect their intentions rather than actual practice. Finally, IOs with more action plan development experience may have different views and adopt different practices than those with less experience. Our findings do not speak to these potential differences. However, it is worth noting that the IOs we interviewed were overall highly experienced. Despite these limitations, the root causes and action plan activities identified by IOs in diverse compliance areas are relevant to noncompliance and integrity violation cases broadly and come from a sizeable sample of 47 IOs. Samples between 12 and 30 are considered suitable for identifying common themes in qualitative research (Corbin and Strauss 2015; Guest, Bunce, and Johnson 2006). Furthermore, while root causes and action plan activities were identified with research leaders in mind, many are applicable to other individuals who may engage in research noncompliance or integrity violations in the conduct research (e.g., postdocs, grad students).

Conclusion

We identified what IOs viewed as common root causes of research noncompliance and integrity violations and common action plan activities that IOs use to address these root causes. The scientific and regulatory communities, including researchers and IOs, would benefit from more in-depth understanding of the complexities and interconnected nature of root causes and various activities that can be implemented in tandem to address these causes. We hope that future efforts will focus on identifying and addressing barriers to developing and implementing robust action plans that effectively address individual, group, and contextual root causes of research noncompliance and integrity violations.

Higher education governing boards play an important role in the state postsecondary education policy arena. As envisioned, these “buffer” organizations work “in the middle” to balance public interests and state goals with the increasingly diverse and evolving priorities of their multiple institutional constituents (McGuiness, 1997, p. 17). In this way, multi-campus systems fulfill dualistic roles as both institutional regulators and institutional protectors. As the former, system governing boards not only hold administrative oversight over their colleges and universities—including through employing presidents/chancellors, approving academic programs and appointments, and allocating resources among institutions— but, as the latter, they also serve as “political intermediaries” between policy actors and their institutions, as well as inter-system referees between institutions themselves, shielding or partially deflecting constituent members from these external and internal pressures (Knott & Payne, 2004; Morgan et al., 2020, p. 3). For institutions, membership in such a system may carry many benefits. As “super-coalitions of sub-coalitions‚” systems draw collective power and resources from their many component parts (Birnbaum, 1988, p. 132). This centralization may provide many benefits, including when seeking additional state or federal support or commanding influence over state policy agenda-setting, but these advantages may also be accompanied by realities that limit the pursuits of individual institutions (Geiger, 2004; Kezar, 2006). In tandem with their duty to advance state goals, promote the public agenda, and act as stewards for their member institutions, system governing boards are also charged with the collective success and preservation of the systems themselves (Berdahl, 1971; Rippner, 2015). That is, accomplishing collective goals sometimes means prioritizing the pursuits of one campus over another or putting the needs of the system before the wants of individual institutions (Knott & Payne, 2004). This collective approach may also include work to inhibit inter-system competition among members, and, at times, leveraging a centralized capacity to cross-subsidize member campuses by redistributing resources from one institution to another or from institutions to the central administration (Richardson et al., 1998). In direct ways, governing boards “regulate and hold universities accountable” through various mechanisms, including by setting tuition and fee rates and approving campus budgets (McGuinness, 1997, p. 12). Beyond these administrative functions, however, Berdahl (1985) argued that the creation of state governing boards was to broadly “protect diversity” by regulating academic drift among institutions (p. 303), suggesting these boards also act as regulators of mission creep to preserve institutional diversity (Morphew & Huisman, 2002). These limits have also been characterized as “mitigating forces against some institutions being able to completely … mimic others,” thus not only regulating institutional aspirations but also constraining the potentially entrepreneurial mechanisms by which they can pursue their priorities (O’Meara, 2007, p. 169). Similarly, regarding inter-system competition, Tandberg (2013) noted that state higher education governing boards may themselves “condition the impacts or effects one actor has on the other,” offering that “were it not for the existence of the boundary-spanning organization or actor, the influence the interacting entities have on each other might be quite different,” suggesting that systems may also help lessen the negative impacts that some members may have on others by directly and indirectly reducing competition for, among others, faculty, students, and other scarce resources (pp. 509–510). This could include providing protection from executive or faculty poaching, duplication of academic programs, monopolization of the state agenda, and more through direct public policymaking, private internal negotiation, and other centralized activities (Tandberg, 2013). As regulators and protectors, system governing boards exercise broad authority to achieve their goals; from maintaining the stratification of institutions within their purview and overseeing individual budgetary decisions, to mediating external political influence and working to ensure institutions do not harm one another or the larger system itself. So, what happens to institutions when the oversight of a system governing board is removed? That is, what happens when multiple campuses from one system are “freed,” independent to pursue altered missions, fully compete, and spend according to individual rather than collective priorities? In this study, I explore these questions using a natural experiment where Tennessee decoupled six public universities from a centralized governing board, granting them autonomy under new and independent governing boards. Specifically, I ask whether this decentralization and rise of a new higher education market organization impacted institutional expenditures in three areas prior works have closely tied to institutional entrepreneurship, competition, and prestige-seeking: executive compensation, faculty salaries, and spending on research activities (Hunt et al., 2019; Marginson, 2004; O’Meara, 2007; Sam & van der Sijde, 2014; Volkwein & Sweitzer, 2006). Understanding these impacts and answering these questions are important for many reasons. To my knowledge, this is the first study to directly consider the effect of increased institutional autonomy on expenditures and one of few to study the effects of a large-scale reorganization of one state’s higher education sector on institutional spending. In this context, I seek to contribute to extant knowledge on the entrepreneurial, competitive, and prestige-seeking behaviors of institutions and to provide useful evidence for policymakers considering the intended (and potentially unintended) consequences of similar governance reorganizations. How institutions use their resources matters for students and states. These decisions can promote or inhibit student success (Gansemer-Topf & Schuh, 2006; Pike et al., 2006), fuel competition within the market (e.g., for faculty and students; Brewer et al., 2009; Ehrenberg, 2003), increase academic and economic stratification within the higher education sector (O’Meara, 2007; Taylor, 2016), and promote inefficiency (e.g., academic program duplication; Dill, 2001). Furthermore, if such a decentralization increased executive compensation, for example, policymakers may not find this to be a prudent use of scarce public resources. Fully exploring these outcomes is necessary to consider the impact of the “protection” and “regulation” functions of system governing boards on individual institutions. While the reorganization of higher education is not new to states (see McGuinness, 2016 for a review), most changes have been piecemeal over time, with few states ever successfully making substantial changes to the complete administration and governance of their system(s) (Bastedo, 2012; McLendon, 2003).1 Thus, this complete removal of a centralized governing board presents a unique opportunity to contribute to extant research in these areas and provide actionable information for policymakers. To accomplish these aims, I begin by providing an overview of the policy landscape and natural experiment I use to assess the effect that greater institutional autonomy has on these expenditures. I then discuss and draw from public administration and institutional theories to frame this study as I consider changes in institutions’ governance structures, motivators in an altered competitive marketplace, and the availability of new mechanisms to pursue altered missions, fully compete, and spend according to individual rather than collective priorities. Next, I review existing literature on state governance reforms and these entrepreneurial, competitive, and prestige-seeking expenditure areas (i.e., executive compensation, faculty salaries, and research) with a focus on institutions’ motivations to alter spending in these areas and the mechanisms by which such deregulation allows them to do so. What follows includes a description of the data and complementary empirical strategies I employ and my results, including a series of associated robustness checks. I conclude with a discussion of the findings and implications of this research for future work and public policy. Tennessee and the FOCUS Act In 2016, Tennessee made a substantial change to its postsecondary sector by decoupling six universities from the Tennessee Board of Regents (TBR) through the Focus on College and University Success (FOCUS) Act.2 At the time, TBR was one of the largest systems of higher education in the nation, governing 46 institutions that served nearly 200,000 students in fall 2015 (Tennessee Higher Education Commission [THEC], 2016a). Describing the rationale for the governance change, the statewide coordinating board noted: “The Act provides greater autonomy for universities in pursuit of innovation and differentiation, while allowing [TBR] to sharpen its attention on technical and community college success. At the core of FOCUS is a belief that increasing the number of Tennesseans with a postsecondary credential demands increased agility on the part of the six TBR universities, with an understanding that this nimbler approach must be deployed within the broader context of the State’s higher education policy agenda” (THEC, 2016b, p. 1). To provide this greater autonomy, the FOCUS Act removed six public universities from TBR and granted them independence under six new and separate governing boards. The FOCUS Act was passed and signed into law during the state’s 2016 legislative session, with the Act taking effect July 1, 2016. While the six universities assumed independence on July 1, 2016, granting first autonomy for the 2016–17 academic/fiscal year, their new governing boards were not installed until the following year’s legislative session.3 Each board held its inaugural meeting beginning in March 2017. Though the FOCUS Act’s hallmark change was a fundamental shift in the state’s postsecondary governance structure, the Act also included other changes, including increasing the regulatory authority of the statewide coordinating board (i.e., THEC) in matters concerning academics, finance, and data, though THEC gained no governing power over any of the state’s public systems or institutions, nor direct oversight of institutional expenditures.

At its core, the FOCUS Act removed oversight of TBR over the six universities and invested this power in six new boards with singular foci on their respective institution. Among these full administrative powers, the FOCUS Act specifically entrusted the new Boards of Trustees to assume the management and governance of each state university, including to “select and employ chief executive officers, confirm [the] appointment of administrative personnel, prescribe curricula, approve operating budgets and set fiscal policies, establish policies and regulations, and assume general responsibility for operation” (TBR, 2016). Each of these duties was previously carried out by TBR, including the selection, evaluation, and compensation of each campus’ president/chancellor; approval of academic programs and the conferral of tenure; administration and review of each campus’ operating budget; and the management of campus’ mission, differentiation, and relations with other system campuses. Upon enactment of the FOCUS Act, institutions began to embody this new freedom. As one salient example of expanded pursuits, the University of Memphis changed its mission statement during the seating of its new board in 2017 to include: “The University of Memphis is a comprehensive metropolitan research university classified by the Carnegie Foundation as Doctoral: Higher Research, but with the goal of reaching Carnegie Very High Research status [emphasis added]” (THEC, 2017, p. 7). In the following year, the university also announced it would alter its expenditures by overturning a TBR policy that limited compensation for part-time instructors. In doing so, the campus noted “one of the first moves by the University of Memphis when we left the TBR system was to boost adjunct pay” (Cook, 2018). In that same year, the regional comprehensive Austin Peay State University (classified as Master’s Colleges & Universities by the Carnegie Classification of Institutions of Higher Education [Carnegie]) announced it would establish its first doctoral program, entering the competitive doctoral arena with the rest of its five FOCUS peers (The Leaf Chronicle, 2017). Finally, in 2020, The Chronicle of Higher Education identified the presidents of East Tennessee State University and the University of Memphis as “movers and shakers,” making the second and third largest jumps, respectively, up the rankings of public university leaders’ pay (Bauman & Elias, 2020). While certain activities of the former system are observable limiters on these behaviors (e.g., the TBR policy referenced by the University of Memphis that limited faculty pay), these immediate changes across many campus domains suggest there were other centralized activities that constrained the FOCUS institutions. It is highly plausible that the advent of an independent governing board provided opportunities for institutions to pursue these individualized missions, alter expenditures, and engage in entrepreneurial ventures in new and intensified ways. Given this “greater autonomy” for the six public universities and the Act’s intended focus on their “pursuit of innovation and differentiation … [with] increased agility” (THEC, 2016b, p. 1), I hypothesize that the removal of centralized oversight from TBR and the installment of individual and independent governing boards provided an opportune environment for the six universities to alter their entrepreneurial, competitive, and prestige- seeking expenditures to higher levels (and via mechanisms) than were previously not possible under centralized control. This is not only exhibited by the above actions of the campuses, but, at a fundamental level, these six universities are no longer members of one postsecondary system, a reality meaning that they must not only increasingly compete in intra- and inter-state markets for faculty, students, and other scarce resources but must also do so under two new realities: they no longer benefit from the protection of a system that concentrates power and shields them from external pressures, but they also no longer have many of the corresponding regulations limiting their entrepreneurial, competitive, prestigeseeking, and other efforts. To leverage this governance change as an opportunity to understand how increased institutional autonomy affects institutional behaviors, I frame the current study in public administration and institutional theories considering the mechanisms and motivations by which institutions might alter their behavior in light of a new governance structure. Conceptual Perspectives In the context of the current study, it is important to acknowledge that this change in governance means many things for the six institutions and the higher education market in the state. Metaphorically, while the “game” remains the same, the players, the coaches, the referee, and the rules have changed. First, these six universities were once part of one team (i.e., TBR) that is now broken into six separate players. Second, the organization of the state higher education sector changed. Each university is now governed by a new board, meaning each answers to a new authority which is also uniquely focused on their success, rather than the success of any system. Third, TBR’s oversight and role as a buffer between institutions has been eliminated. While the statewide coordinating agency remains present, THEC holds no governing power and has a limited ability to “condition the impacts or effects one [university] has on the other” (Tandberg, 2013, p. 509). Finally, while each board and university is still expected to advance state goals and promote the public agenda, each board’s focused oversight of one institution rather than many means the “innovation and differentiation” each institution pursues, the “autonomy” by which they pursue it, and the “agility” or “nimbler” mechanisms by which they do so can vary from those previously controlled by TBR (THEC, 2016b, p. 1). These realities characterize a new higher education market in Tennessee which may not only allow the six universities to alter their entrepreneurial, competitive, and prestige-seeking expenditures to higher levels and via altered mechanisms than before but may also encourage such behaviors given a greater need to compete in an increasingly complex intra-state market. To consider these changing players and changing contexts, I draw from principal-agent theory and notions of academic capitalism, entrepreneurialism, and revenue theory of costs to help frame the current study and describe how the removal of centralized control—and the regulation and protection it provided—may incent changes in institutional behavior. Principal-agent theory concerns itself with the roles and relationship between two parties, a leader (the “principal”) and a follower (the “agent”), wherein the principal contracts the agent to perform duties on the principal’s behalf (Jensen & Meckling, 1976; Moe, 1984). This framework has been widely applied to the field of higher education to consider issues of governance and policy (see Yallew et al., 2018 for a systematic review). Given that both public and private institutions of higher education are relatively autonomous, and the fact that public principals (e.g., the state or large boards) often have multiple agents, prior work has contextualized the role of the principal as exercising a variety of “oversight” or “monitoring” roles for a “semi-autonomous” agent (Lane, 2007, p. 622). In the current context, principal-agent theory is useful given the fact that each institution (i.e., agent) is now directed by new boards (i.e., principal), and that the focus of these new principals (i.e., on one institution) varies from that of the previous one (i.e., TBR’s focus on the system at large). Given these changes, the agent should not only be freer to pursue an altered mission and fully compete—including the ability to alter spending in areas that support entrepreneurialism, competition, and the maximization of prestige or other resources—but should also have greater flexibility of tools with which to do so. This is particularly likely given the fact that agreements between principals and agents rest upon a belief that action of the agent will improve the status of the principal relative to others (Alchian & Demsetz, 1972; Lane & Kivisto, 2008). In this light, independent boards with foci on singular institutions are likely to permit or encourage institutional behaviors which benefit the institution’s pursuits and are unlikely to exercise Tandberg’s (2013) conditioning behaviors to mitigate their effects on others, unlike that of a centralized board. Indeed, prior studies have shown that a simple change in the principal alone has been enough to incent changes in agents’ behaviors (Kwiek, 2021; Lane & Kivisto, 2008; Liefner, 2003). For these reasons, I hypothesize that the six universities will spend differently under decentralized versus centralized control. Given that the six universities now operate in a heightened competitive market absent any system-level referee (i.e., the presence of new, autonomous agents similarly seeking students, faculty, prestige, state appropriations, and other scarce resources in ways that are unconditioned by a system governing board), the notions of academic capitalism, entrepreneurialism, and revenue theory of costs are also important to consider. Academic capitalism and entrepreneurialism posit that institutions have necessarily adopted broad market- like behaviors in their pursuit of power, prestige, and resources (Slaughter & Leslie, 1997; Slaughter & Rhoades, 2004), while Bowen’s (1980) revenue theory of costs suggests organizations will maximize revenues to the fullest extent possible to fund infinitely evolving operations (Archibald & Feldman, 2008; Cheslock et al., 2016; Kimball, 2014). In this context, faced with greater competition and equipped with greater autonomy, the six FOCUS universities are thus likely to engage in the new market in a different manner than they previously did as they employ “nimbler approach[es]” with heightened “agility” (THEC, 2016b, p. 1). Fundamentally, given institutions’ tendencies to seek new resources for endless pursuits, and systems’ operations to regulate, limit academic drift, and condition behavior one has on another, the removal of TBR was also the removal of a barrier to set and pursue such endless goals—and the removal of an external body to define the ways in which those goals could be pursued. In all, the six universities and their boards now not only operate as separate autonomous agents, heightening their need to compete (and the severity of a failure to do), but they may now also pursue expanded aims through innovative and entrepreneurial efforts. Indeed, it could be expected that a market characterized by heightened competition and complexity alone would lead institutions to increasingly adopt such market-like behaviors that advance their competitive advantage and to focus expenditures on areas that support this pursuit. Discussing this likelihood, Eckel (2008) argued that, “when faced with choices of where to invest scarce time and resources, institutions may yield to those priorities that will position them well in the competitive arena” (p. 184). For these reasons, I hypothesize that the six universities will not only spend differently under decentralized versus centralized control, but that these altered expenditures will be focused in areas that can provide a competitive advantage, including ones that may have been previously limited or conditioned by a centralized authority. In this study, I ask whether this decentralization and altered market impacted institutional expenditures. While many actions may improve institutions’ position within the higher education market, governing boards and campus administrators have the most immediate control over expenditures, which ultimately hold the potential to impact shortand long-term institutional outcomes (Ryan, 2004; Webber & Ehrenberg, 2010). Prior works have consistently identified three areas of expenditures commonly associated with institutional entrepreneurship, competition, and prestige-seeking: executive compensation, faculty salaries, and spending on research activities. As discussed above, these are also areas that have been directly identified or exhibited by the FOCUS universities as areas constrained by a centralized system. In what follows, I introduce prior work on the antecedents to state higher education governance reforms, review the extant literature on these expenditure areas, and discuss how the FOCUS Act enables and motivates institutions to increase expenditures in pursuit of a heightened competitive advantage and resource acquisition. Literature Review As noted, many states have made piecemeal changes to their higher education governing structures, but few have made large-scale reorganizations via a single act in recent history (McGuinness, 2016). While this study leverages a unique opportunity to estimate impacts of one such large change, it is founded upon a rich body of work that considers the antecedents to these reforms. McLendon et al. (2007) reviewed many of these studies and conducted an analysis examining determinants of 22 governance reforms enacted from 1985–2000. The authors found that unified legislative control and larger shares of Republicans were strong predictors of organizational changes, and Tandberg and Anderson (2012) reinforced these findings by suggesting that governance changes are driven by “economic pressures [and] efforts to reinvent government,” including state responses to “growth in the size, complexity, and cost of higher education” (p. 565). Indeed, a host of work has identified politics and political maneuvering as an impetus for many higher education governance reorganizations (MacTaggart, 1996; McLendon, 2003; McLendon & Ness, 2003), and, beyond these political forces, Sponsler (2010) noted that states broadly also adopt policies given “learning, competition, coercion, and socialization” (p. 47). In Tennessee’s context, the 2016 state legislature had a Republican supermajority (85% of the senate and 74% of the house), and Republican Bill Haslam was still in his second term as governor (Council of State Governments, 2016), suggesting that the political environment was primed for a governance reform. Similarly, the state cited a need for campuses to pursue “innovation and differentiation” with “increased agility” and “nimbler approach[es]” in service of the “State’s higher education policy agenda” (THEC, 2016b, p. 1). These terms pull from much of the “new public management” and “quasi-public corporation” language used in prior reforms to dually suggest that such a change can (1) increase the public benefit of institutions while also (2) leveraging many private-sector efficiencies or practices (McLendon et al., 2007, p. 648). In all, such a reform to Tennessee’s higher education governance structure could have been expected, particularly given the suite of policies introduced by the same governor and legislature (e.g., Drive to 55, Tennessee Promise). However, while such reforms can appear rational or efficient, existing work finds generally null impacts of them on state higher education performance (Conner & Rabovsky, 2011; Volkwein, 1986; Volkwein & Tandberg, 2008), and others find such reorganizations may in fact leave states with increased costs given requirements to carry out and sustain such reforms in practice (Heller, 2003). As the FOCUS universities embark under a new governance structure that increases autonomy and removes the inter-campus “buffer” previously fulfilled by TBR, competition for faculty, students, and other scarce resources (including state appropriations) is likely to increase. Thus, institutions are likely to accentuate their focus on increasing power, resources, and prestige in ways that are objectively and subjectively associated with the accumulation of these commodities (Dill, 2001; Eckel, 2008). Prior studies have considered many “striving behaviors” of institutions (O’Meara, 2007, p. 122), including how institutions’ strategic priorities to maximize prestige, acquire additional resources, and improve their competitive advantage can traverse all areas of campus life, such as within academic programs (Maguad, 2018; Morphew, 2000), student recruitment and engagement (Bock et al., 2014; Szekeres, 2010; Zilvinskis & Rocconi, 2018), faculty hiring and evaluation (Gardner & Veliz, 2014; Johnson, 2017), institutional identity (Miller, 2019; Rusch & Wilbur, 2007), tuition and fee pricing (Allen & Shen, 1999), campus facilities (Griffith, 1999; McClure, 2019), and more. A key strand of this prior work has also considered institutions’ administrative expenditures or resource allocation (e.g., McClure & Titus, 2018; Morphew & Baker, 2004; Santos, 2007), and three areas of expenditures that have been commonly associated with institutional entrepreneurship, competition, and prestige-seeking include executive compensation, faculty salaries, and spending on research activities (Hunt et al., 2019; Marginson, 2004; O’Meara, 2007; Sam & van der Sijde, 2014; Volkwein & Sweitzer, 2006). While no work to date has documented the impacts of a centralized governing authority on these expenditure areas, a robust body of work has linked these spending categories to institutional “striving” for additional resources and competitive advantages (McClure & Titus, 2018; Ehrenberg et al., 2007; Leslie et al., 2012; O’Meara, 2007, p. 122). Monks (2007) and Tang et al. (1996) found presidential pay rates to be associated with institutional rankings and prestige, and Pfeffer and Ross (1988) argued that presidential pay rates “reflect the practices and premises that pervade the organization” given the heightened visibility of a president (p. 79). If institutions seek to reflect power and success, highly compensating a chief executive has shown to be a viable strategy. Pfeffer and Ross (1988) underscored this signaling effect of high-wage presidents (i.e., having a “valuable” president), noting that a clear measure of leader effectiveness is the organizational accumulation of resources. Given that prestigious institutions have accumulated, among others, wealth and power, they may signal the success of their president with higher compensation packages. Other works have also observed that presidents of higher-ranked and better-resourced institutions (i.e., land-grant and R1 [very high research activity] institutions) earn more on average than the chief executives of less prestigious institutions (Bartlett & Sorokina, 2005; Ehrenberg et al., 2001; Hunt et al., 2019). In the Tennessee context, the newly independent boards now have full control over the compensation of their president/chancellor, which was previously set by TBR. As institutional complexity and competition increase in the deregulated market, the role of a president/chancellor may more closely resemble that of a private Chief Executive Officer, providing upward pressure on their compensation while more closely tying that compensation to institutional success (Cheng, 2014; Huang & Chen, 2013). Indeed, this appears to have occurred in Tennessee shortly following the FOCUS Act as shown by large descriptive changes in the compensation of two FOCUS universities’ presidents as reported by The Chronicle of Higher Education. After observing that institutions invest their scarce resources on areas that yield competitive advantages, Eckel (2008) noted that institutions are also rewarded by “focusing on…star faculty [and] externally funded research” (p. 184). Characteristics of the faculty are strong predictors of institutional productivity and resource acquisition, including the number of citations, level of external funding, educational experiences (i.e., highest-degree and doctoral-granting institution), and publication counts (Volkwein & Sweitzer, 2006). The payment of the faculty is no exception. In fact, Melguizo and Strober (2007) observed that faculty members are financially rewarded for succeeding at many activities that enhance institutional quality and prestige, including by attracting other faculty, securing large research grants, recruiting strong undergraduate and graduate students, and securing donors who want to be associated with institutions that are “winners” (p. 638). Competition for such faculty has led to a “silent crisis” for public institutions as they work to recruit and retain star faculty within their fiscal constraints (Alexander, 2001). Furthermore, in a mechanical sense, faculty salaries increase prestige and rankings—both strong predictors of competition and subsequent resource acquisition (Bastedo & Bowman, 2010; Monks & Ehrenberg, 1999; Pike, 2004)—by directly influencing U.S. News & World Report scores (Morse et al., 2019). In Tennessee, the new governing boards now not only set academic policies and confer tenure, but also directly approve institutional operating budgets, including salary changes that were, as noted by the University of Memphis, limited by TBR. Under increased competition for high-quality faculty that can in turn confer prestige and acquire resources for the institution, it could be expected that the six FOCUS universities may feel pressure to increase faculty compensation, particularly if faculty are expected to increase productivity in light of an institution’s heightened research agenda (e.g., the University of Memphis’ new Carnegie goal or Austin Peay State University’s new doctoral program). This upward pressure on salaries is also likely given the fact that TBR no longer serves as a “buffer” between these institutions to regulate activities like poaching or academic program duplication, increasing the need for institutions to compete for and retain faculty. Similar to achieving a higher ranking in part by increasing faculty pay, achieving Carnegie R1 (very high research activity) status through increased spending on research activities can also bring prestige and heighten institutions’ competitive advantage. McClure and Titus (2018) note that achieving R1 status, an explicit goal of the newly independent University of Memphis, “represents an honorific in higher education that carries certain benefits” (p. 969), where others have suggested such an honorific can have positive effects on applications and selectivity, media and press coverage, and subsequent financial earnings (Bowman & Bastedo, 2009; Hearn & Rosinger, 2014; O’Meara, 2007). In Tennessee specifically, even if institutions are not pursuing R1 status, any spending on research activities is directly tied to the state’s outcomes-based funding model that distributes 100% of state appropriations (THEC, 2020). Here, institutions’ spending on research, service, and sponsored programs represents 10–15% of the outcomes upon which the state bases its annual allocation of over $1 billion in funding. Thus, increasing expenditures on research represents an opportune area to not only invest in activities associated with other positive outcomes, but also an area that is likely to yield a strong return on investment. The newly formed FOCUS governing boards now control institutional operating budgets and set institutional priorities, including the pursuit of augmented research goals, thus streamlining their ability and agility to increase expenditures in this area highly associated with prestige, rankings, and the accumulation of additional campus resources. While I hypothesize that the removal of TBR’s oversight from the six FOCUS universities may enable and encourage altered expenditure patterns in light of heightened competition for students, faculty, and other scarce resources, it should be noted that not all activities of states necessarily reduce institutions’ abilities to compete. For example, many states play increasing roles in the university research enterprise, providing support by way of funding general enrollment subsidies to research universities; making direct, large-scale investments in research as a mechanism to increase economic development; passing state tax incentives for public–private research partnerships or research collaborations; or taking similar actions that can centrally support institutions’ research competitiveness (Feldman & Desrochers, 2003; Feller, 2004; Hearn & Lacy, 2009; Mowery et al., 2001; Plosila, 2004).

Similarly, on the faculty front, many states now support “eminent scholars” programs with taxpayer funds—programs that provide funding for the recruitment of accomplished scientists to fulfill endowed chairs and professorships (Hearn et al., 2013). Activities like these suggest that states can use their centralized authority and resources to support the pursuits of individual institutions. However, it is also important to recognize that such programs may exist alongside the presence of a centralized governing board to mediate or “condition” this competition and work to ensure universities leverage these resources to compete regionally or nationally—rather than intra-state against one another (Tandberg, 2013, p. 509). Without such an actor, like TBR in Tennessee, such state programs may have varied impacts. On one hand, they may continue to support institutions striving to achieve higher levels of research activity by providing them with necessary capital. On the other, they may exacerbate existing inequalities between institutions by increasing the competitive advantage of already-well-resourced institutions—particularly for programs that leverage stateuniversity matching funds—which could ultimately increase inefficiency and disadvantage emerging universities in the new competitive arena. In all, state oversight does not limit institutions’ competitive abilities per se, but such programs may be best suited in the presence of a centralized governing board to mitigate any unintended consequences. Drawing upon these conceptual foundations and prior works, I leverage the FOCUS Act as a natural experiment to assess how increased this institutional autonomy impacted executive compensation, faculty salaries, and spending on research activities. In what follows, I review the data for the study and the complementary causal inference techniques I employ to answer this question. In doing so, I not only seek to advance the nascent literature on how state governance structures influence institutional behaviors but also seek to provide useful evidence for policymakers considering the intended (and potentially unintended) consequences of similar governance reorganizations. Data Data for this study are drawn from three primary sources: The Chronicle of Higher Education’s Executive Compensation at Public and Private Colleges survey, THEC administrative records on president/chancellor compensation and benefits, and the U.S. Department of Education’s Integrated Postsecondary Education Data System (IPEDS). I derive executive compensation information for the six public universities in Tennessee and several comparison institutions from the former two sources, while I pull faculty salaries, research expenditures, and institution-level covariate controls from the latter. The Chronicle maintains a website and dataset with results from its annual executive compensation survey, which covered the compensation of chief executives at more than 600 private institutions and nearly 270 public institution and systems in 2019 (Bauman et al., 2020). These data were graciously provided by The Chronicle for this project. This is, to my knowledge, the largest source of executive compensation information on higher education leaders since IPEDS does not report these data. Given that the FOCUS Act applied only to public institutions in Tennessee, I focus only on those public institutions included in The Chronicle’s survey to assemble an initial comparison group. After excluding any leaders that represent systems (e.g., University of Texas or University of North Carolina systems), I collected the base pay rates for each remaining institution from academic/ fiscal years 2010–11 through 2016–17 so long as the institution was represented in the survey for at least four years (n = 160). For years after 2016–17, The Chronicle moved its reporting window to focus on calendar year (e.g., 2018) rather than academic/fiscal year (e.g., 2017–18) compensation to mirror that of its private college data. I collected the same campus-by-year compensation for these leaders in the 2018 and 2019 calendar years (the most recent available). For institutions with mid-year changes in leadership (e.g., two or more presidents/chancellors in a given year), I averaged each pay rate to arrive at one executive compensation rate by campus by year, and for any campus reporting less than the full nine years of data (n = 25), I linearly interpolated these values. The final Chronicle dataset represents the base compensation rates of presidents/chancellors of 160 public institutions (not including any in Tennessee), covering academic/fiscal years 2010–11 through 2016–17, calendar year 2018 (spanning academic/fiscal years 2017–18 and 2018–19), and calendar year 2019 (spanning 2018–19 and 2019–20). Among the 6 institutions impacted by Tennessee’s FOCUS Act, only 5 have ever reported compensation information to The Chronicle, and only 3 have done so at least half of the survey’s duration. Upon request and agreement, however, THEC provided executive compensation information for all public institutions in the state from academic/fiscal year 2009–10 through 2019–20, including current salary levels for the chief executive. These are public data collected annually by THEC as part of institutional operating budgets. None of the six universities experienced a change in president/chancellor following the FOCUS Act, and none had mid-year leadership changes. Given these are administrative records, there were also no missing values in the current salary variable, which is akin to The Chronicle’s base pay rate which excludes bonuses or other compensation.5 To bridge The Chronicle’s 2018 calendar year reporting change, I averaged the THEC-reported salaries for 2017–18 and 2018–19 given that the 2018 calendar year spans both periods and did the same for 2019 (averaging 2018–19 and 2019–20).6 In all, the combined compensation file includes base salaries for institutional chief executives from 2010–11 through 2018–19/2019 for 160 public institutions across the nation and 9 public institutions in Tennessee (the six former TBR universities and three other public universities in the state). Finally, I construct an institution-level panel from IPEDS covering my remaining two outcome variables of interest: the average, 9-month-equivalent salary of a full professor and current year total expenditures on research activities, as well as covariate controls from academic/fiscal year 2010–11 through 2018–19. These data cover the 169 public universities available within The Chronicle and THEC datasets. Control variables include fall full-time equivalent (FTE) enrollment; the six-year Bachelor’s degree graduation rate; the percent of first-time full-time students who receive Pell grants; the percent of students who are a racial minority7; the current tuition and mandatory fee rate; the institution’s total annual operating revenues, endowment balance, and admissions rate; the Carnegie classification designation by year (defined as 1 for doctoral universities and 0 otherwise); and the 75th percentile score of the ACT composite for the admitted class. Prior research guided the selection of these controls related to entrepreneurialism, competition, and prestige, and previous works have found these factors to be predictive of executive compensation, faculty salaries, and research expenditures (Bartlett & Sorokina, 2005; Boudreau et al., 1997; Ehrenberg et al., 2001; Hearn, 1999; Hunt et al., 2019; Johnson et al., 1987; Monks, 2007; Riggs et al., 1986). For each of these outcomes and controls, I again linearly interpolate missing values; seven institutions were dropped for not reporting at least two years of data for an outcome or covariates to be interpolated. For one treated institution in Tennessee, neither the admission rate nor the ACT composite were ever reported, so these values were imputed with the grand mean.

In all, the final dataset covers the expenditures of interest and multiple covariate controls from 2010–11 through 2018–19/2019 for 162 public universities. This represents 6 years of observations prior to FOCUS and 3 years after (the most recent available). All financial figures were adjusted for inflation using the 2019 Consumer Price Index. Descriptive statistics for the three outcomes of interest and 10 covariate controls across the six FOCUS universities, the 156 comparison institutions, and the total sample for the 2015–16 academic/ fiscal year (immediately prior to the FOCUS Act) are presented in Table 1. Methods In this study, I leverage two complementary strategies to estimate causal impacts of the FOCUS Act on institutions’ executive compensation, faculty salary, and research activity spending. I first implement a traditional difference-in-differences design and then supplement this strategy with a synthetic control approach. Both strategies yield valid causal inferences under the assumptions discussed below, yet the latter allows the difference-indifferences’ parallel-trend assumption to be empirically relaxed. Difference‑in‑Differences Given that TBR’s oversight was removed beginning July 1, 2016 for all FOCUS universities, I first employ a difference-in-differences (DID) empirical strategy to estimate the causal effect of this change on institutions’ expenditures. DID is a common quasi-experimental technique that exploits across-unit and inter-temporal variation and is a preferred estimation strategy when assessing the effects of fixed-time policy adoptions because its ability to addresses concerns of selection and omitted variable bias (Angrist & Pischke). Here, I seek to compare the six universities in Tennessee (i.e., treatment) to a counterfactual group of all other colleges that were not affected by this policy change (i.e., control) by examining differences in the outcomes of interest both before and after the 2016 FOCUS Act. Formally, I estimate: where yit is one of three prestige-seeking outcomes of interest for institution i in year t ; FOCUSit is a binary indicator identifying treatment (i.e., the FOCUS Act), which takes the value of 1 in 2016–17 and later for the six formerly-TBR universities or 0 otherwise; and ¬ it is a vector of institution-specific, time-varying controls described earlier. For the 2018 (or 2019) calendar year compensation rate, I use 2017–18 (or 2018–19) academic/ fiscal year predictors (the most recent available). The specification is also conditioned on unit ( ¬i ) and year ( ¬t ) fixed effects to absorb unobserved unit-specific, time-invariant and across-unit, year-specific factors. Here, ¬ is the parameter of interest, or the causal effect of the FOCUS Act on an institution’s given yit expenditures. Each model is weighted by FTE enrollment. I estimate heteroscedastic-robust standard errors and cluster at the state level (i.e., highest level and the level of treatment) to account for serial correlation in outcomes (Abadie et al., 2017; Bertrand et al., 2004; Cameron & Miller, 2015). I also log each outcome and financial control given skewed distributions.

The primary assumption embedded in DID is that the treatment and control group would exhibit a constant difference in outcome trends in absence of treatment (Angrist & Pischke, 2009, 2015). While this parallel-trends assumption is untestable in the potential-outcomes framework (Rubin, 2005), I plot outcome trends for the six FOCUS universities and controls in the top row of Fig. 1. In aggregate, the parallel-trends assumption is plausible in the pre-treatment period (2010–11 through 2015–16) for each outcome of interest, where both groups follow similar outcome trajectories. In the post-treatment period (2016–17 and later), increases in each outcome are noted for the treatment group which may be attributable to impacts from the FOCUS Act. Though parallel trends are strongly plausible here, it is possible that the six universities in Tennessee systematically varied on these outcomes in such a way that significant estimates could be produced even in the absence of treatment. If these universities varied from the counterfactual group in the pre-treatment period, then any estimated post-treatment differences could be due to these systematic differences rather than being representative of a FOCUS Act treatment effect. To assess this possibility, I conduct an event study analysis to test for effects pre-treatment years (i.e., similar to multiple placebo tests) and to assess the magnitude of differences in the post-treatment period. This further assesses the appropriateness of DID in this setting. Formally, I estimate: where the specification is the same as Eq. 1, but I now interact a dummy indicator equal to 1 for the FOCUS institutions ( FOCUSi ; 0 otherwise) with each year factor ( ¬t ) from 2010–11 through 2018–19, omitting the year immediately prior to the FOCUS Act (2015–16) as reference. For underlying model assumptions to be met, I expect significant differences to be absent across the 2010–11 through 2014–15 horizon, or for effects to be in the opposite direction as hypothesized (i.e., where Tennessee institutions expended less funds on each outcome of interest), but for significant effects to be estimated following the governance change (2016–17 and later). Results of this specification are plotted in the bottom row of Fig. 1, where each annual estimate (compared to 2015–16) is plotted and bounded by a 95% confidence interval. The president/chancellor compensation outcomes meet these assumptions, with no statistically significant outcome differences between the FOCUS institutions and the controls in the pre-treatment period, suggesting DID is an appropriate strategy. The research expenditures outcome also meets this assumption for all but 1 pre-treatment year (2013–14), where the 95% interval barely crosses 0 (0.0018). This is the only pre-treatment year where significant impacts would be expected, which is (1) not nominally similar to the magnitude of post-treatment effects estimated (i.e., 0.0655 in 2013–14 compared to 0.1704 in 2016–17) and (2) not suggestive evidence of sustained systematic differences between the treatment and control groups. For the average professor salary outcome, however, there are sustained and time-variant differences between the six FOCUS institutions and the control group (i.e., 2012–13 and 2013–14), suggesting a violation of the parallel-trends assumption for this outcome. While DID appears to be an appropriate strategy for assessing impacts on president/chancellor compensation and research expenditures—and all eventstudy plots point to large and statistically significant increases following the FOCUS Act for each outcome of interest—these differences in the average professor salary urge caution when interpreting their results and motivate the complementary use of an identification strategy which overcomes this issue by empirically relaxing this parallel-trends assumption: synthetic control. Synthetic Control Pioneered by Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015), synthetic control methods (like DID estimators) are used to compare outcomes among treatment and control groups before and after a policy change, with observed outcome differences following a policy’s implementation attributable to the effect of the policy. Referred to as a generalized extension of DID, GSCM uses all available treatment–control comparison points and weights units to create a comparison group that is nearly identical to the treatment group on outcomes in the pre-treatment period, allowing DID’s strict parallel-trends assumption to be relaxed (Cunningham, 2021). By generating a synthetic control unit whose outcomes mirror that of the treatment group conditioned on observable characteristics prior to the policy change (i.e., by shrinking pre-treatment differences between treatment and control groups toward zero), concerns regarding the selection of an optimal comparison group are reduced, and the synthetic unit can be considered a suitable comparison given its statistically indistinguishable difference from the treatment group (Abadie et al., 2010, 2015). The application of synthetic control methods to education is emerging yet remains underutilized (Jaquette et al., 2018; Rubin & González Canché, 2019; Ward & Ost, 2021). Unlike a DID application, where researchers guide the selection of a control group which could violate its parallel-trends assumption, synthetic control methods create a suitable counterfactual from all available control units such that Here, in the pre-treatment period ( t = 0 ), the average outcome ( Y ) for all universities not impacted by the FOCUS Act (i.e., controls) are weighted to equal those of a FOCUS institution ( yit ). By considering each control unit i ’s outcome as a linear function of observable covariates in the pre-treatment period, an optimization algorithm identifies a weight w for each control unit i such that the wi optimal weight for control unit i ensures yFOCUS i,t=0 and ΣwiY Control t=0 are as mathematically close as possible. Knowing that the mean outcome difference between treatment and control units in the pre-treatment period is as mathematically as close to zero as possible, ΣwiY Control t=0 becomes a suitable counterfactual, eliminating parallel-trend concerns.8 The generalized synthetic control method (GSCM) follows the same intuition but allows for multiple treatment units (i.e., six FOCUS universities) by aggregating separate synthetic units for each treated unit with a linear interactive fixed effects model (Krief et al., 2016, 2018, 2020; Xu, 2017; Xu & Liu, 2018, 2020). When predicting outcomes, interactive fixed effects models interact unit-specific intercepts (“factor loadings”) with timevarying coefficients (“factors”) such that where yit is the outcome for unit i in year t , conditioned on unit (¬i) and year (¬t) fixed effects; FOCUSit is a time-varying treatment indicator equal to 1 for the FOCUS universities in the post-treatment period (0 otherwise); and ¬ it is the vector of controls. Here, ¬ ¬ i are unit factor loadings interacted with time-varying factors ft , and ¬it is the heterogenous average treatment effect on the treated (ATT) estimate for unit i at time t . The number of factors and their factor loadings are derived by an optimization procedure that performs the equivalent function of the wi weights in Eq. (3) to optimally weight control units (Xu, 2017)9. Aggregating these ¬it impacts for N units in treatment group T produces the overall ATT of exposure to the FOCUS Act, shown by which is the average annual difference between treatment units [Y(1)] and their synthetic control units not exposed to FOCUS [Y(0)] in the post-treatment period. I implement GSCM by estimating Eq. (4) above for each expenditure outcome of interest and allow the optimal weighting algorithm to pull from all available control institutions in the Chronicle universe to construct a suitable counterfactual unit.10 The GSCM equivalent to the DID parallel-trend plots are presented in Fig. 2. These figures show the superior control of GSCM over parallel trends and exhibit the respective outcome deviations between the six FOCUS universities (“Treated Average”) and their aggregated synthetic control unit (“Estimated Y(0) Average”). Here, the optimal weighting process achieved strong alignment between the treatment and synthetic control unit(s) evidenced by the minimal- to-zero differences between each line in the pre-treatment period, suggesting GSCM is an appropriate strategy in this context and should complement the DID estimates. In the post-treatment (shaded) period, like the DID event-study plots, these figures also show increases in the FOCUS universities’ institutional expenditures in each of these categories (observed by gaps between their raw outcomes and the weighted counterfactuals).

Results

Difference-in-differences and synthetic control estimates of the effect of institutional autonomy on each expenditure category are presented in Table 2 by outcome and estimation strategy. For the executive compensation outcome, the DID model with full covariate controls and both institution and year fixed effects suggests the six FOCUS universities increased the base pay rate of their president/chancellor by approximately 6.17% ¬ exp(0.0598) − 1× 100. Given a baseline (cumulative pre-treatment) mean of approximately $305,130, this increase is equivalent to an increase of over $18,830. The GSCM estimator produces a similar finding, suggesting institutions increased executive compensation by roughly 6.21%, or $18,950. For the professor salary outcome, both estimates also point to consistent increases in faculty compensation, ranging from 1.59% ($1,440) in the DID model to 2.19% ($1,980) in the GSCM model. As noted, while the DID estimator likely fails to meet the parallel-trends assumption for this outcome, the qualitatively similar GSCM estimate provides additional and plausible evidence of increases in faculty compensation. Finally, concerning research expenditures, both models again point to consistent and statistically significant increases in spending, ranging from approximately to $1.59 M (9.41%) in the GSCM model to $2.0 M (11.8%) in the DID model. In all, these results provide consistent evidence to suggest that the removal of a centralized governing board from the six universities and installment of individual and independent boards provided an opportunity and environment for institutions to increase spending in on entrepreneurial, competitive, prestige-seeking activities.

Robustness

This study is not without notable limitations in design or data quality. Here, I detail two main points and conduct a series of robustness checks to further consider underlying assumptions of the analysis and to test the stability of the main findings to altered specifications.

Treatment Period

A primary limitation could concern identification of the post-treatment period. As noted, the FOCUS Act was passed during the 2016 legislative session, taking effect July 1, 2016. This meant the six public universities were autonomous beginning in the 2016–17 academic/fiscal year, and the primary models in Eqs. (1) and (4) treat 2016–17 as the first year of treatment. However, as previously noted, these institutions’ new governing boards did not sit for the first time until March 2017. While institutions had autonomy from TBR during the 2016–17 academic/fiscal year, it is possible the full effect of a new governance structure could not be detected until the 2017–18 academic/fiscal year (or, for compensation, the 2018 calendar year) when the new governing board could consider or alter institutions’ expenditures, such as setting new compensation rates for the president/chancellor. To consider this possibility, I re-estimate the main models but lag the treatment indicator so that the pre-treatment period covers academic/fiscal years 2010–11 through 2016–17 and the post-treatment period is defined as 2017–18/2018 and 2018–19/2019 only. Comparing these two specifications allows me to consider the earliest possible effects of the FOCUS Act on institutions’ expenditures (2016–17 and later) and possible effects after full implementation (2017–18/2018 and later). Results from this lagged treatment specification are presented in Table 3. As expected, outcome estimates are larger and more statistically significant, suggesting that either the seating of the independent governing board could have propelled spending or that the full effect of the FOCUS Act would not be experienced until the academic/fiscal (or calendar) year after the full governance transition was complete. Estimates here are again consistent across strategies and suggest statistically significant increases in executive compensation, professor salaries, and research expenditures. Results suggests the six FOCUS universities increased executive compensation by approximately 8.15–8.86% ($25,180–27,370), increased professor salaries by 2.80–2.89% ($2,540–2,620), and increased research spending by 9.79–13.17% ($1.67–2.24 M). These lagged increases are also supported by the raw and weighted outcome plots shown in Figs. 1 (row 2) and 2. Taken together with the primary 2016–17 academic/fiscal year treatment specification, these results provide additional and consistent evidence of increases in institutions’ president/chancellor compensation, faculty salaries, and research expenditures following the governance change.

Executive Compensation

A second concern is a data limitation brought about by The Chronicle’s change from academic/ fiscal year compensation reporting to calendar year reporting in 2018. As shown in Fig. 1’s parallel trends, there is an increase for all groups in the 2017–18/2018 year, likely due to this mechanical change. While this is a concern, it is important to note that this change applied to all institutions, and the means-based DID and GSCM estimators are agnostic to this difference given that all institutions’ compensation rates were shifted upward. That is, there is no evidence to suggest that institutions in one state or system would benefit over another by such a reporting change. It is also important to note here that using the THEC salary data for all public institutions in Tennessee (which is used in each model) represents a conservative salary estimate given that THEC data match almost exactly to The Chronicle in 2016–17 yet underestimate the 2018 calendar year salary (as discussed in the Data section). Detecting any significant increase between these years for the treated institutions should therefore be regarded as particularly substantial. While I cannot formally overcome this data limitation, I do test the robustness of these findings in two ways. First, I shorten the post-period window to only include 2016–17 when there was no change in reporting for institutions. While the previous section (Treatment Period) may suggest this is too early to detect full effects of the FOCUS Act, any significant changes in the executive compensation outcome detected would be based only on a constantly of the full effect given that it only considers the first year of treatment. Second, I limit the treated sample to those three FOCUS universities that responded to The Chronicle’s compensation survey for at least four years and replace their THEC-computed 2018 and 2019 calendar year salaries with The Chronicle’s reported salary, allowing their mechanical change to exactly mirror that of the other comparison institutions. Results for these specifications are presented in Tables 4 and 5. For the shortened 2016–17 post-treatment window (Table 4), estimated impacts are smaller yet less-precisely-estimated than the main models, suggesting an increase in president/chancellor salaries for the six universities of 1.60% ($4,880) in the DID model and 3.07% ($9,370) in the GSCM model, though neither are statistically significant. This reduced precision is likely due to the limitations imposed with only one year of post-treatment data and the prior-observed and possible lagged impacts of the FOCUS Act. However, for the subset of three universities who reported to The Chronicle, where I observe a constant transition from academic/fiscal to calendar year reporting, the models suggest larger effects given the conservative nature of the THEC data (Table 5). Estimates range from an approximately $75,450 (23.58%) increase in the DID model to $76,340 (23.86%) increase in the GSCM model. Each of these specifications again provides further confidence in the main findings of a significant increase in executive compensation following institutional autonomy. In all, results from the main DID and GSCM models and those across several robustness checks provide early yet consistent results suggesting that the removal of a centralized governing board from the six universities and installment of individual and independent boards provided an opportunity and environment for institutions to increase spending on executive compensation, professor salaries, and research. These robustness tests also underscore the likelihood that the full impacts of such a governance change may not be felt until a few years after the transition.

Discussion

As regulators and protectors, system governing boards exercise broad authority to administratively oversee their colleges and universities and to serve as intermediaries between these institutions, one another, and the external environment (Knott & Payne, 2004; McGuinness, 1997; Morgan et al., 2020). While these protective activities shield institutions to a large degree from external political influence and undue intra-system competition, the corresponding regulatory activities may limit the ability of individual campuses to pursue altered missions, fully compete, and spend according to individual rather than collective priorities (Berdahl, 1971; Geiger, 2004; Kezar, 2006; Rippner, 2015). These regulations may manifest through direct public policymaking, private internal negotiation, and other centralized activities (Tandberg, 2013). Yet despite research to date on state and system higher education governance structures, no work has considered how the removal of a centralized governing board impacts subsequent institutional behavior. To explore this question, I leveraged a natural experiment where Tennessee “freed” six of its public universities in 2016 by removing the oversight of a centralized board and investing this power in six new boards with a singular focus on their own institution. Given a novel opportunity to assess how institutions respond to a new intra-state market characterized by deregulation and increased competition for faculty, students, and other scarce resources given newly autonomous actors, I focused on changes in three institutional expenditure areas closely tied to entrepreneurship, competition, and prestige-seeking: executive compensation, faculty salaries, and spending on research activities. With a primary goal to provide “greater autonomy” for the six universities to pursue “innovation and differentiation” by employing “nimbler approach[es]” with heightened “agility” (THEC, 2016b, p. 1), the FOCUS Act yielded a new higher education market in Tennessee. At a fundamental level, six universities were no longer members of one postsecondary system, a reality meaning that they must not only increasingly compete but must also do so under two additional realities: they no longer benefit from the protection of a system that concentrated power and shielded them from external pressures, but they are also no longer subject to many of the corresponding regulations that limited their entrepreneurial, competitive, prestige-seeking, and other efforts. Thus, considering the control of a new governing board, I drew from principal-agent theory to hypothesize that these institutions should not only be freer to pursue an altered mission and fully compete—including the ability to alter spending in areas that support entrepreneurialism, competition, and the maximization of prestige or other resources—but should also have greater flexibility of tools with which to do so. In this light, I believe independent boards with foci on singular institutions are likely to permit or encourage institutional behaviors which benefit the institution’s pursuits and are unlikely to exercise Tandberg’s (2013) conditioning behaviors to mitigate their effects on others—unlike that of a centralized board. Thus, the removal of TBR was not only the removal of a barrier to set and pursue new goals, but it was also the removal of an external body that defined the ways in which those goals could be pursued. While certain activities of the former system were observable limiters on these behaviors (e.g., the TBR policy referenced by the University of Memphis that limited faculty pay), the immediate changes in executive compensation, the establishment of new doctoral programs, and the alteration of faculty pay policies across many campuses suggest there were other centralized activities that previously constrained the FOCUS institutions. As the target area for these increased entrepreneurial, competitive, and prestige-seeking behaviors, I drew from institutional notions of academic capitalism, entrepreneurialism, and revenue theory of costs, as well as from prior works on institutional expenditures, to identify three areas of spending that institutions were likely to alter. While many actions may improve institutions’ position within the higher education market, governing boards and campus administrators have the most immediate control over expenditures (Ryan, 2004; Webber & Ehrenberg, 2010). In this scenario, I hypothesized that the six universities would not only spend differently under decentralized versus centralized control, but that these altered expenditures would be focused in areas that can provide a competitive advantage by maximizing power, prestige, and the accumulation of other resources—and that may have been previously limited or conditioned by a centralized authority. Using complementary difference-in-differences and synthetic control approaches, I found robust evidence suggesting that newly independent universities increased the salary of their president/ chancellor by approximately 6.2% (or $19,000), increased the average full professor salary by 2.2% (nearly $2,000), and increased research expenditures by an average of 12% (or $2 million). These findings support the primary hypotheses and suggest that the deregulation of Tennessee’s higher education market produced a structure that not only allowed the six universities to alter their expenditures to higher levels than before but one that also encouraged such behaviors given a greater need to compete in an increasingly complex intra-state market. Whether expenditures or activities in these areas were directly or indirectly prohibited by TBR, it is undoubtable that the FOCUS Act signaled an opportunity for institutions to alter spending and resulted in large and significant increases in their expenditures across these areas. The results of this study have important implications for public policy and future research. Fundamentally, I asked whether institutions increased expenditures in three areas following a governance change, and robust evidence suggests the answer is yes. A host of prior works have considered institutions’ administrative expenditures or resource allocation in pursuit of power, prestige, and resources in these and other areas of campus operations (e.g., Hunt et al., 2019; Marginson, 2004; McClure & Titus, 2018; Morphew & Baker, 2004; O’Meara, 2007; Sam & van der Sijde, 2014; Santos, 2007; Volkwein & Sweitzer, 2006). While the FOCUS universities still have governing boards, it would be irrational for a board to prohibit behaviors of the institution that, ceteris paribus, benefit its ability to compete and accumulate additional tangible and intangible resources. Being home to institutions (and, particularly, public ones) that increasingly spend, compete, and move up the prestige ladder can be two sided for states. On one hand, stronger institutions can increase enrollment levels, increase the price elasticity of demand for college (i.e., allowing institutions to raise tuition, freeing demand for state appropriations), increase other revenues for higher education (e.g., philanthropy or external research support), or impact a variety of state-related outcomes (e.g., in and out migration or workforce development). On the other hand, supporting increasingly complex and successful institutions, particularly those developed through increased expenditures, requires resources to maintain such quality and standing (McClure & Titus, 2018; Morphew & Baker, 2004). Likewise, this may introduce additional issues brought about through increased competition that may, in tandem, promote inefficiency (e.g., academic program duplication or increased demand on state appropriations to support research activities; Dill, 2001; Volkwein & Sweitzer, 2006). Indeed, policymakers and the public alike may not find increased executive compensation to be a prudent use of scarce public resources. Policymakers should carefully consider these as intended (and potentially unintended) consequences of such a governance reorganization insofar as it may affect institutions’ spending patterns. Furthermore, policymakers should also be aware that some state-sponsored programs to support institutional competition (e.g., research investments, eminent scholars) could have varied impacts in the absence of a centralized governing board. Such activities could be used to help mitigate negative consequences of increased competition (e.g., by supporting developing research universities) but could also exacerbate existing resource inequities between universities if not closely monitored. To further aid these considerations, future work in this area is needed. To my knowledge, this is the first study to consider the effect of increased institutional autonomy on expenditures and one of few to study the effects of a large-scale reorganization of a state’s higher education sector on spending behaviors. In this light, my work builds upon a strong foundation of prior literature on governance and institutional expenditures while working to connect the two given a novel opportunity to assess how institutions respond to increased autonomy to pursue augmented missions, fully compete, and alter spending. In doing so, my findings reinforce existing conceptual foundations regarding the nature and function of system governing boards and extends these prior works to more fully consider how system governing boards’ roles as regulators and protectors condition these institutional behaviors. Future work should consider other possible impacts of such a reorganization on institutional behavior, including possible changes to admissions, tuition and financial aid, athletics and auxiliary enterprises, spending on student services, and more. While the current study narrowly focused on outcomes that have been empirically associated with entrepreneurialism, competition, and prestige, future studies should consider other expenditure activities or behaviors that could promote or inhibit student success (Gansemer-Topf & Schuh, 2006; Pike et al., 2006), fuel greater competition (Brewer et al., 2009; Ehrenberg, 2003), or increase academic and economic stratification within the higher education sector (O’Meara, 2007; Taylor, 2016). Indeed, while such higher education governance reforms can appear rational or efficient, existing work finds generally null impacts of them on subsequent state higher education performance but does find that such reorganizations can actually leave states with increased costs to support and sustain the governance changes in practice (Conner & Rabovsky, 2011; Heller, 2003; Volkwein, 1986; Volkwein & Tandberg, 2008). The costs and benefits of these possible intended (and unintended) consequences must be further explored by future research and fully considered by state policymakers.