

Research excellence indicators: time to reimagine the ‘making of’?

Federico Ferretti^{1,*}, Ângela Guimarães Pereira¹, Dániel Vértessy¹ and Sjoerd Hardeman²

¹European Commission, Joint Research Centre, Via E. Fermi 2749, Ispra 21027, Italy and ²RaboResearch Nederland, Rabobank, The Netherlands

*Corresponding author. Email: federico.ferretti@ec.europa.eu

Abstract

In the current parlance of evidence-based policy, indicators are increasingly called upon to inform policymakers, including in the research and innovation domain. However, few studies have scrutinized how such indicators come about in practice. We take as an example the development of an indicator by the European Commission, the Research Excellence in Science & Technology indicator. First, we outline tensions related to defining and measuring research excellence for policy using the notion of ‘essentially contested concept’. Second, we explore the construction and use of the aforementioned indicator through in-depth interviews with relevant actors and the co-production of indicators, that is the interplay of their making vis-à-vis academic practices and policy expectations. We find that although many respondents in our study feel uncomfortable with the current usage of notions of excellence as indicator of quality of research practices, few alternatives are suggested. We identify a number of challenges which may contribute to the debate of indicator development, suggesting that the making of current indicators for research policy in the EU may be in need of serious review.

Key words: Research Excellence indicators; Science and Technology indicators; Quantification; Essentially contested concepts; Social construction of indicators

1. Introduction

When it comes to research policy, excellence is on top of the agenda. Yet, the meaning attributed to the notion of excellence differs markedly among both academics and policymakers alike.

There is an extensive scholarly debate around the breadth and depth of the meaning of excellence, its capacity to provide quantitative assessments of research activities and its potential to support policy choices. Yet, there is a considerable agreement that it strongly influences the conduction of science. The contentedness of the excellence concept can be derived from the discomfort it has evoked among scholars, leading some even to plea for an altogether rejection of the concept (Stilgoe 2015). The discomfort with the concept is higher whenever proposals are made to measure it. The critique of measuring excellence follows two lines. One is technical and emphasises the need for methodological rigour. While in principle not denying the need for and the possibility of designing science and technology indicators, this line of criticism stresses the shortcomings of methodological approaches used up until now (Grupp and Mogee 2004; Grupp and Schubert 2010). The other critique is more philosophical and, while not denying the theoretical and political relevance of excellence, it takes issue with the use of current metrics in

assessing it (Weingart 2005; Martin 2011; Sørensen et al. 2015). Notwithstanding these criticisms though, and especially given the period of science professionalization where policymaking finds itself in (Elzinga 2012), these same metrics are frequently called upon to legitimate policy interventions (Wilsdon et al. 2015).

In addition, highly reflected shortcomings in the existing mechanisms for science’s quality control system, undermine trust in assessment practices around scientific excellence—in other words, if the peer review system is in crisis, what research outcomes are evaluated as excellent? (See Martin 2013; Sarewitz 2015; Saltelli and Funtowicz 2017.)

The aspiration for an ‘evidence-based society’ (Smith 1996) requests that policy makers and alike, especially those operating at the level of transnational governmental organisations, rely on information on the current state of research to identify policy priorities, or to allocate funds. Indicators are typically proposed as tools catering this need (Saltelli et al. 2011). A main issue holds, however: how to come up with indicators of research excellence in the face of its often controversial underpinnings, as well as their situated nature?

At the Joint Research Centre of the European Commission, we have been actively involved in the design and construction of a country-level indicator of excellence, the Research Excellence

Science & Technology indicator (RES&T) offered and used by the European Commission (cf. [European Commission 2014](#); [Hardeman et al. 2013](#)). Hence we are in a unique position to critically reflect upon challenges of quantifying research excellence for policy purposes.

Here we adopt the notion of *essentially contested concept* as our theoretical work horse ([Gallie 1955](#); [Collier et al. 2006](#)) to discuss why the usefulness of research excellence for policy purposes is a subject of contention and what this means for its quantification. Essentially contested concepts are concepts ‘the proper use of which inevitably involves endless disputes about their proper uses on the part of their users’ ([Gallie 1955](#): 169).

The work presented in this article revolves around two questions which evolved with the learning through the empirical material: First, we examine whether research excellence can be ‘institutionalised’ in the form of stable research excellence indicators, from the vantage point of Gallie’s notion of ‘essentially contested concept’. Second, whether the re-negotiation of meanings of research excellence that underpin current indicators revolves around the articulation of different imaginaries of excellence displayed by different actors. These initial questions were reframed with the progressive understanding of the authors that the focus in the practices were certainly relevant but larger questions emerged, such as whether ‘excellence’ alone was indeed the relevant descriptor to evaluate quality of research in the EU. Hence, this discussion is also offered vis-à-vis our findings throughout the research process.

The article starts by looking into the notion of excellence and its function as a proxy for scientific quality using the notion of essentially contested concept as well as elements of tension around its conceptualization (Section 2) as reported in the literature. It proceeds with describing briefly the development of the indicator that we are taking as an example to respond to the research questions described earlier. The second part of the article explains the methodology applied (Section 3) and the outcomes (Section 4) of the empirical research carried out to inform this article, which consisted of a number of in-depth interviews with relevant actors, that is developers of the RES&T indicator, EU policymakers, and academics. The interviews aimed at exploring meanings, challenges, and ways to reimagine the processes behind indicators development. In those interviews, we explore ‘re-imagination’ as a space for our interviewees to reflect further and discuss alternatives to current research indicators frameworks. These are offered in a discussion (Section 5) of current challenges to reimagine an indicator to qualify quality in science.

2. Quantifying research excellence for policy purposes

2.1 Measuring and quantifying indicators-for-policy

The appeal of numbers is especially compelling to bureaucratic officials who lack a mandate of popular election or divine right; scientific objectivity thus provides an answer to a moral demand for impartiality and fairness; is a way of making decisions without seeming to decide. (T. M. Porter 1995)

Indicators seek to put into numbers phenomena that are hard to measure ([Boulanger 2014](#); [Porter 2015](#)). Therewith, measuring is something else than quantifying ([Desrosieres 2015](#)): while measuring is about putting into numbers something that already exists, quantifying is about putting into numbers something that requires

an interpretative act. Indicators are often exemplary of quantifications. They are desirable because they offer narratives to simplify complex phenomena and therewith attempt to render them comprehensible ([Espeland 2015](#)). Such simplifications are especially appealing whenever information is called for by policymakers operating at a distance from the real contexts that is, the actual purpose of their policy action. Simplification means that someone decides which aspects of complex phenomena are stripped away while others are taken on board. The (knowledge and values) grounds for that operation are not always visible. The risk is that, in stripping away some aspects (and focusing on others), a distorted view on the phenomenon of interest may arise, with potentially severe consequences for policy decisions derived from them. Lacking the opportunity to gather detailed information on each and every aspect of a phenomenon of concern, policymakers are nevertheless drawn to indicators offering them the information needed in the form of summary accounts ([Porter 2015](#)).

Constructing an indicator on research excellence typically involves activities of quantification as research excellence has no physical substance in itself. For an indicator on research excellence to come into existence one first needs a meaning and understanding about what ‘research excellence’ is about before one can even start assigning numbers to the concept ([Barré 2001](#)). We find that the notion of ‘co-production’ ([Jasanoff 2004](#)) is relevant as it makes visible that indicators are not developed in a vacuum but respond and simultaneously normalise scientific practice and policy expectations.

2.2 Research excellence as an essentially contested concept

Research excellence could be straightforwardly defined as *going beyond a superior standard in research* ([Tijssen 2003](#)). However, straightforward and intuitively appealing as this definition may seem, it merely shifts the issue of defining what is meant by research excellence towards what counts as ‘a superior standard in research’. For one thing, it remains unclear what should be counted as research to begin with, as well as how standards of superiority should be set, on which account and by whom. Overall, the notion of research excellence is potentially much more controversial than it might seem at first. In fact, whenever it comes to articulating what should count as excellent research and why this is so, scientific communities systematically strive for coming to an agreement ([Lamont 2009](#)).

One way to conceive of research excellence then, is to think of it as an *essentially contested concept*. The notion of *essentially contested concept* was first introduced by [Gallie \(1955\)](#) to describe cases, that is ideas or phenomena that are widely appraised but controversial at the same time. In substantiating his view, [Gallie \(1955\)](#) listed five properties of essentially contested concepts (see also: [Collier et al. 2006](#)). Essentially contested concepts are (1) *appraisive*, (2) internally complex, (3) describable in multiple ways, (4) inherently open, and (5) recognized reciprocally among different parties ([Gallie 1955](#)). Due to their complex, open and value-laden nature, essentially contested concepts cannot be defined in a single-best, fixed, and objective way from the outset. Hence, they are likely to produce endless debates on their interpretation and implications.

Research excellence might well serve as an instance of an essentially contested concept. First, research excellence, by its very appeal to superior standards, evokes a general sense of worth and, therewith, *shareability*. Although one can argue about its exact definition and the implications that such definitions could have, it is hard to be against excellence altogether ([Stilgoe 2015](#)). Second, research excellence is

likely to be internally complex as it pertains to elements of the research enterprise that need not be additive in straightforward ways.

For example, research excellence can be about process as well as outcomes, whereby the former need not automatically transform into the latter (Merton 1973). Third, it follows that research excellence can be described in multiple ways: while some might simply speak of research excellence with reference to science's peer review system (Tijssen 2003), others prefer to broaden the notion of research excellence beyond its internal value system to include science's wider societal impact as well (Stilgoe 2015). Fourth, what counts as excellent research now might not necessary count as excellent research in the future, and any definition of research excellence might well be subject to revision. Finally, the fact that one can have a different view on what research excellence is or should be, is agreed upon by proponents of different definitions. Ultimately, proponents of a particular notion of research excellence *could* or *could not* be aware of alternative interpretations.

Recently, Sir Keith Burnett (2016) argued that a mechanical vision of academia is driving 'mechanical and conventional ways we think about "excellence"'. We measure a community of scholars in forms of published papers and league tables' (Burnett 2016). Hence, what counts as excellence is entertained by the imagination of some about what 'excellent research' is; but what, political, social, and ethical commitments are built into the adopted notion and the choice of what needs to be quantified?

2.3 Quantifying research excellence for policy purposes: critical issues

Following the previous discussion, if one acknowledges research excellence as an essentially contested concept, the construction of indicators faces difficulties, which start with the mere act of attempting quantification, that is agreeing on a shared meaning of research excellence. In the 1970s Merton (1973: 433–435) introduced three questions that need to be addressed to come to terms with the notion of research excellence (see also: Sorensen et al. 2015).

First, what is the basic unit of analysis to which research excellence pertains? Merton (1973) suggested that this could be everything ranging from a discovery, a paper, a painting, a building, a book, a sculpture, a symphony, a person's life work, or an oeuvre. There is both a temporal, as well as a socio-spatial dimension to the identification of a unit for research excellence. Temporal in the sense that research excellence does not need to be attributable to a specific point in time only but might span across larger time periods. Also, though not so much discussed by Merton (1973), a unit for research excellence also has a socio-spatial dimension. Research excellence might pertain to objects (books, papers, sculptures, etc.) or people. When it comes to the latter a major issue holds to whom excellence can be attributed (individuals, groups, organisations, territories) and how to draw appropriate boundaries among them (cf. Hardeman 2013). Expanding or restricting a unit's range in time and/or space effects the quantification of research excellence accordingly.

Second, what qualities of research excellence are to be judged? Beyond the identification of an appropriate unit of analysis, the second issue raised by Merton (1973) points out several concerns. One is about the domain of research itself. As with disputes about science and non-science (Gieryn 1983), demarcating research from non-research is more easily said than done. Yet, to attribute excellence to research, such *boundary work* needs to be done nevertheless. Should research excellence, as in the article *Republic of Science* (Polanyi 1962) be judged according to its own criteria? Or should research,

in line with Weinberg's (1962) emphasis on external criteria, be judged according to its contribution to society at large? To the same extent of setting the unit of excellence, setting the qualities in one way (and not another) produce certainly different outcomes for the policies derived therefrom. That said, focusing on a particular notion of excellence (i.e. using a particular set of qualities) might crowd out other—in principle equally valid—qualities (Rafols et al. 2012; Sorensen et al. 2015).

Third, who shall judge? For example, a researcher working in a public lab might have a whole different idea on what counts as excellent research than one working in a private lab. This brings Stilgoe (2015) to argue that "Excellence" tells us nothing about how important the science is and everything about who decides'. It is undoubtedly of eminent importance to determine the goals and interests that excellence serves. Likewise, and in line with Funtowicz and Ravetz's (1990) focus on fit-for-purpose to describe the quality of a process or product, the quality of an indicator of research excellence crucially depends on its use. One concern here is that research excellence indicators might set the standard of research practices that do not conform to the underlying concept of excellence they seek to achieve (Hicks 2012; Sorensen et al. 2015). For example, in Australia, in seeking to achieve excellence, the explicit focus on publication output indeed increased the number of papers produced but left the issue of the actual worth of those 'unaddressed' papers (Butler 2003). Interestingly, in 2009 a new *excellence framework* came into existence in Australia to replace the former *quality framework*. While the latter made use of a one-size-fits-all model, the new excellence based one presents a matrix approach in which entire sets of indicators, as well as the experts' reviews coexist as measures of quality. Again, any definition of research excellence and its implications for quantification need to be positioned against the background of the goals and interests it serves.

2.4 The construction of the Research Excellence Indicator (RES&T) at the Joint Research Centre

The development of the Research Excellence Indicator (RES&T) at the Joint Research Centre of the European Commission (JRC) inspired this research. Its history, developments, and our privileged position of proximity to its developments motivate the basis from which we departed to conduct our inquiries.

In 2011, an expert group on the measurement of innovation set up by the European Commission's Directorate-General Research and Innovation (DG-RTD) was requested 'to reflect on the indicators which are the most relevant to describe the progress to excellence of European research' (Barré et al. 2011: 3). At that point the whole notion of excellence was said to be 'in a rather fuzzy state' (Barré et al. 2011: 3). To overcome the conceptual confusion surrounding research excellence and to come up with a short list of indicators capable of grasping research excellence, the expert group proceeded in four steps. First, they defined and described types of activities eligible for being called excellent. Second, a set of potential indicators were identified. Third, from this set of potential indicators a short list of (actually available) indicators was recommended. And fourth, a process for interpreting research excellence as a whole at the level of countries was proposed.

This was followed by Vertesy and Tarantola (2012) proposing ways to aggregate the set of indicators identified by the expert group into a single composite index measuring research excellence. The index closely resembled the theoretical framework offered by the expert group while aiming for statistical soundness at the same time.

Presented at a workshop organised in Ispra (Italy) during fall 2012 by the European Commission and attended by both policy-makers and academic scholars, the newly proposed composite indicator met with fierce criticism. A first critique raised was that the proposed composite indicator mixed up both inputs and outputs while research excellence, according to the critiques, should be about research outputs only. Whereas the outcomes of research and innovation activities are fundamentally uncertain, the nature and magnitude of research and innovation inputs say little to nothing about their outputs. A second critique raised during the workshop was that some of the indicators used, while certainly pertaining to research, need not say much about their excellent content. Focusing on outputs only, would largely exclude other dimensions that could refer to any kind of input (e.g. gross investment in R&D) or any kind of process organizing the translation of inputs into outputs (e.g. university–industry collaborations).

Taking these critiques on board, the research excellence indicator was further refined towards the finalization of the 2013 report (Hardeman et al. 2013). First, the scope of the indicator was made explicit by limiting it to research in science and technology only. Second, following upon the critique strongly to distinguish inputs from outputs, it was put clear which among underlying indicators were primarily focused on outputs. Given that the underlying indicators were not available for all countries, the rankings presented in the 2013 Innovation Union Competitiveness Report was based on a single composite indicator aggregating either three (non-ERA countries) or four (ERA countries) underlying indicators (European Commission 2013).

In a subsequent report aimed at refining the indicator, Hardeman and Vertesy (2015) addressed a number of methodological choices, some of which were also pointed out by Sørensen et al. (2015). These concerned the scope of coverage in terms of the number and kind of countries and the range of (consecutive) years, the variables included (both numerators and denominators), and the choice of weighting and aggregating components. The sensitivity and uncertainty analyses highlighted that some of the methodological choices were more influential than others. While these findings highlighted the importance of normative choices, such normative debates materialized only within a limited arena.

Based on our research and experience with the RES&T, we will discuss whether careful reconsideration of the processes by which these types of indicators are developed and applied is needed.

3. Methodology

A qualitative social research methodology is adopted to gain insights from different actors' vantage point on concepts, challenges, and practices that sustain the quantification of research excellence.

A series of in-depth interviews was carried out by two of the authors of this paper between March and May 2016. A first set of interviews was conducted with five people directly involved in the construction of the RES&T indicator from the policy and research spheres, or people that were identified through our review of relevant literature. This was followed by a second set of interviews partially suggested by the interviewees in the first set. Hence, eleven telephone semi-structured in-depth interviews were conducted with experts, scholars, and users concerned with research indicators and metrics.

This was followed by a second set of interviews (six participants), partially suggested by the interviewees in the first set.

This second set was thus composed of senior managers and scholars of research centres with departments on scientometrics and bibliometrics, as well as policymakers.

Hence, the eleven interviewees included people that were either involved in different phases of the RES&T indicator development or were professionally close (in research or policy areas) to the topic of research indicators and metrics. Awareness of the RES&T indicator constituted a preferable requirement. The eleven telephone semi-structured in-depth interviews conducted with the experts, scholars, and users of indicators may seem numerically little; however, this pool offered relevant insights to shed light on the practices of research evaluation in the EU. So, the interviewees were the relevant actors for our work.

We performed coding as suggested by Clarke (2003) as soon as data were available and such an approach allowed us setting more focus on some aspects of the research that emerged as particularly important. The accuracy of our interpretations was checked through multiple blind comparisons of the coding generated by the authors of this paper. Often our codes have also explicitly been verified with the interviewees to check potential misalignments in the representativeness of our interpretations.

For the purposes of our research the interviewees were grouped into the following categories:

1. RES&T indicator developers (hereafter referred to as 'developers')
The three interviewees of this group were all somehow involved in the design and implementation of the RES&T indicator. Among them, two senior and one retired researchers, all of them active in the areas of innovation and statistics. Given that we knew two of the interviewees before the interview, we paid particular attention to the influence of the interviewer–interviewee identities at the moment of data analysis, along the recommendations of Gunasekara (2007).
2. Policy analysts (hereafter referred to as 'users')
This group was composed of four senior experts, who are users of research indicators. They are active as policymakers at the European Commission; they all have been involved in various expert groups and at least two of them have also published own research.
3. Practitioners and scholars in relevant fields for our endeavour concerned with science and technology indicators and active at different levels in their conceptualization, use and construction (hereafter referred to as 'practitioners')
This group was composed of four scholars (one senior researcher, one full professor, one department director, and one scientific journal editor in chief) who critically study statistical indicators.

A guiding questionnaire was prepared and tailored for each of the interviewees, according to their expertise and/or relationship with the making of the RES&T indicator. Each interview included three different momentums:

- Insights into meanings of excellence.
- Critical overview of current metrics in general and processes and backstage dynamics in the development of the RES&T indicator (applicable if the interviewee was personally involved).
- Reimagination of ways to assess and assure the quality of processes of indicators development, taking stock of transformations of knowledge production, knowledge governance, and policy needs (new narratives).

All interviews, which were in average one hour long, have been transcribed and data analysis was conducted according the principles of grounded theory, particularly as presented in Charmaz (2006). The analysis of these interviews consisted of highlighting potential common viewpoints, identifying similar themes and patterns around the topics discussed with the interviewees, which will be discussed in the next sections of this article.

4. Meanings, metrics, processes, and reimagination

In this section, we attempt to make sense of the issues raised by our interviewees, summarising the main recurrent elements of the three main axes that were at the core of the questionnaire structure: (1) meanings of 'excellence', (2) challenges backstage processes of developing and using research excellence indicators, (3) ways to reimagine the process of indicator development to promote better quality frameworks to assess scientific quality.

4.1 On meanings of research excellence

Many of us are persuaded that we know what we mean by excellence and would prefer not to be asked to explain. We act as though we believe that close inspection of the idea of excellence will cause it to dissolve into nothing. (Merton 1973: 422)

Our starting question to all interviewees was 'please, define research excellence'. To this question, our interviewees found themselves rather unprepared, which could suggest that either this expression is taken for granted and not in need of reflection or—as the literature review shows—no shared definition seems to exist, to which, in the end, our interviewees largely agree. Such *unpreparedness* seems somehow paradoxical as it implies an assumption that the definition of excellence is stable, in no need for reflection, whereas our interviewees' responses seem to suggest rather the contrary. Excellence is referred to as 'hard to define', 'complex', 'ambiguous', 'dynamic', 'dangerous', 'tricky' as well as, a 'contextual' and 'actor-dependent' notion. The quotes below reflect different vantage points, indicating some agreement on its multidimensional, contested, distributed, situated, and contextual nature:

[...] this is a **dangerous concept**, because you have different starting positions. Developer 3

Clearly, excellence is **multi-dimensional**. Secondly, excellence ought to be considered in **dynamic terms**, and therefore excellence is also dynamics, movement and progress which can be labelled excellent. Third, excellence is not a natural notion in the absolute, **but it is relative to objectives**. Therefore, we immediately enter into a more **complex notion of excellence** I would say, which of course the policy makers do not like because it is more complicated. Developer 2

[...] you need to see the concept of research excellence **from different perspectives**. For universities it might mean one thing, for a private company it might mean something completely different. Developer 1

You could say that excellence is an **emergent property of the system** and not so much an individual attribute of particular people or groups. Stakeholder 1

The quotes suggest agreement among the interviewees that research excellence is a multidimensional, complex, and value-laden concept

which link well with the notion of essentially contested concept introduced earlier. While some experts simply think of highly cited publications as the main ingredient for a quantification of excellence, others tend to problematize the notion of excellence once they are invited to carefully reflect upon it, getting away from the initial *official* viewpoint. Indeed, the *lack of consensus* about meanings of excellence is highlighted by different interviewees, and, not surprisingly, seem to be a rather important issue at the level of institutional users and developers, who described it as an unavoidable limitation. For example:

It is **extremely difficult to have a consensus** and [therefore] it is impossible to have a perfect indicator. User 2

I do see that there was no clear understanding of the concept [of excellence] [since] the Lisbon agenda. This partly explains why [a] high level panel was brought together [by DG RTD], [whose] task was to define it and they gave a very broad definition, but I **would not identify it as the Commission's view**. Developer 3

The way users and developers responded to this lack of consensus seems to be different though. Developers, on the one hand, do not seem to take any definition of research excellence for granted. It seems that, as a way out of the idea that research excellence constitutes an essentially contested concept, developers stick to a rather abstract notion of research excellence: specific dimensions, aggregation methods, and weights are not spelled out in detail. For example, when asked to define excellence, one developer responded:

I would say there is a **natural or obvious standard language meaning**, which is being in the first ranks of competition. Excellence is coming first. Now, we know that **such a simple definition is not very relevant [for the task of indicators making]**. Developer 2

The more concrete, and perhaps more relevant decisions are therewith avoided, as it is immediately acknowledged that research excellence constitutes an essentially contested concept. Users, on the other hand, seem to take for granted established definitions much easier. Here, one interviewee simply referred to the legal basis of Horizon 2020 in defining excellence:

I think I would **stick to the definition of the legal basis**: what is important is that the research that is funded is the top research. How is this defined? In general, it is what is talented, looking for future solutions, preparing for the next generation of science and technology, being able to make breakthroughs in society.¹ User 3

What both developers and users share is their insistence on the need for quantification of research excellence, albeit for different reasons. From the user-perspective, the call for a research excellence indicator seems to be grounded in a desire for evidence-based policy (EBP) making.

To our question on whether excellence is the right concept for assessing quality in science, interviewees responded saying that the call for EBP all costs surely plays a fundamental role in the mechanisms of promotion of excellence measures and therefrom indicators development:

There is a huge debate on what the real impact of that is in investment and we **need to have a more scientific and evidence-based approach to measure this impact**, both to justify the expense and the impact of reform, but also to better allocate spending. User 2

Notwithstanding the difficulty involved in operationalizing a notion of excellence towards indicators, what comes upfront is that no single agreed-upon solution is to be expected from academia when it comes to defining excellence for quantification purposes.

This seems to be acknowledged by one of the developers, commenting on the composition of the high-level expert panel that:

You have a bunch of researchers who have a very different understanding of what research excellence would be, and some were selected for this high level panel. I am not aware of any reasoning why specific researchers or professors were selected while others were not. **I am sure that if there was a different group, there would have been a different outcome**, but this is a tricky thing. Developer 3

As such, similar considerations seem to confirm that the processes behind indicators development, such as the involvement of certain academic communities, potentially influence further conceptualisations of research excellence. These aspects will be discussed in the last section of this article.

4.2 Metrics and processes of research excellence

... the whole indicator activity is a social process, a socio-political process; it is not just a technical process. Developer 2

Indicators and metrics respond and correspond to social and political needs and are not mere technical processes, and this is made visible by different types of tensions identified by our interviewees.

First, the process of quantification of research excellence requires an agreement on its definition. Any definition is neither definitive nor stable, not least because of its contextual dependencies. In the above section, it emerged that *what* needs to be quantified is substantially contested. However, our interviews show that other at least equally contested dimensions exist, such as methodological (quantification practices), social (involved actors), normative (scientific and policy practices).

In the remainder of this section, we explore through our interviews the production of indicators vis-a-vis their processes and outcomes.

4.2.1 Normativity: who is involved in the design of an indicator?

Indicators clearly require political choices to be made. What needs to be quantified and who decides remains an important question. The normativity aspects remit always to definition issues, social actors of concern and institutional dependencies.

The observation of one of the practitioners resonates with Jack Stillgoe's provocation that 'excellence tells us nothing about how important the science is and everything about who decides'.²

Who decides what excellence is? Is it policy makers, is it the citizen, is it the researchers themselves? Is it absolute or does it depend on the conditions? Does it depend on the discipline? Does it depend on the kind of institution concerned? You see what I mean. Developer 2

A practitioner suggests that the level of satisfaction, and therefore acceptance, of an indicator is primarily defined by its usage:

Who will decide when an indicator is good enough and to what extent? [...] The short answer is the users, the people out there who use indicators and also whose careers are affected by indicators, they decide whether it's good enough. Practitioner 3

These quotes raise different questions related to what we here call 'normativity' and ideas of co-production, both in terms of indicators development and usage: first, what are power relations between the actors involved and how can they influence the

processes behind indicators? Second, to what extent can these kinds of quantification be deemed unsatisfactory and, ultimately rejected and by whom? Third, in the idiom of co-production, how does research excellence metrics influence research practices in both mainstream knowledge production systems and other emerging systems of knowledge production (namely what is designated as 'DIY science', 'citizen science', 'the maker movement', etc.)?

4.2.2 Inescapable simplifications?

Simplification seems to be an inescapable avenue in any attempt to represent complex concepts with just one number; as it implies inclusion and exclusion of dimensions, it begs the question of responsibility and accountability. In the end of the day, complex systems of knowledge production are evaluated through very limited information. Although we do not want to expand this discussion herein, it is important to point out that when using these scientific tools in realms that will have major implications on systems of knowledge production and governance, the 'who' and 'to whom' necessarily need careful consideration.

At some point, you need to reduce the complexity of reality. Otherwise you cannot move on. We tend to be in favour of something. The problem is that we have far too many possibilities for indicators [...]. **In general, we need to take decisions on the basis of a limited amount of information.** Practitioner 4

What is the limitation of the index? These were the main issues and dimensions that it was not able to address. I do not know what the most problematic things were. I have seen many questions, which address for instance the choice of a certain indicator, data or denominator and the exclusion or inclusion of an index. I do not know which ones were more important than the others. We ran a number of sensitivity tests, which showed that some of the choices had a more substantial impact on country rankings. You could put the ranks upside down. Developer 3

Different interviewees deem that quantification practices ought to be robust to deviations due to different theoretical assumptions therefrom when specific variables, time periods, weights, and aggregation schemes are varied.

With regards to the RES&T, one user remarked purposeful correction as an issue of major concern for the quantification of research excellence:

Part of [the] alignment [of the RES&T indicator] led to counter-intuitive results, like a very low performance from Denmark, and then we reinstated the previous definition because it led to slightly better results. The definition of results is also important for political acceptance. User 4

As reported in the literature and also emerged throughout our interviews, excellence has been contended as the relevant policy concept to tackle major challenges of measuring impacts of science in society. The two quotes below stress the importance of aiming for indicators that go beyond the mere scientific outputs, suggesting that frameworks of assessment should also encompass other critical issues related to process (e.g. research ethics):

It is OK to measure publications, but not just the number. For instance, also how a certain topic has been developed in a specific domain or has been taken into a wider issue, or any more specific issues, these needs to be tracked as well. User 3

Let's imagine two research groups: one does not do animal testing, and obtain mediocre results, the other does animal testing

and have better results and more publications. **How those two very different ethical approaches can be accounted?** We should correct excellence by ethics! Developer 1

Our material illustrates several facets of different types of reductionism: first, the loss of multidimensionality as an inevitable consequence of reducing complexity; second, rankings following from indicators sometimes work as drivers and specifications for the production of the indicators themselves; finally, volatility in the results is expected to become an issue of major concern specifically along ever-changing systems of knowledge production (see e.g. Hessels and van Lente 2008).

4.2.3 Backstage negotiations

Indicators largely depend on negotiations among actors seeking to implement their own vision and interest. From such a view, research indicators redefine reputation and prioritise funding. This process is depicted as an embedded and inevitable dynamic within indicators production:

[When developing an indicator] you will always have a negotiation process. You will always discuss ‘what you will do in this case’; ‘you would not include that’ or ‘you would do that’; ‘this does not cover it all’. **You will always have imperfect and to a certain extent wrong data in whatever indicator you will have.** User 1

[Developers] mainly do methodological work. The political decisions on the indicator are taken a bit higher up. Practitioner 3

Many politicians have a very poor view of what actually goes into knowledge production. This is what we have experienced in Europe, The Netherlands and the UK. Give me one number and one A4 with a half page summary and I can take decisions. We need to have some condensation and summarisation, and you cannot expect politicians to deal with all the complexities. At the same time, they must be aware that too poor a view of what knowledge production is, kills the chicken that lays the eggs. Practitioner 1

These quotes seem to suggest that there are ‘clear’ separate roles for those who participate in the production of the indicator and those who are empowered to decide what the final product may look like. In the case of the development of the RES&T indicator, the process of revision and validation of the indicator included a workshop organised by EC policymakers, in which developers and academics were invited to review the indicator’s proposed theoretical framework. The publication of the feasibility study by Barré et al. (2011) was the main input of this workshop; one of the developers that we interviewed remarked the following:

I find it interesting that [at the workshop] also policymakers had their own idea of what it [the indicator] **should** be. Developer 3

In other words, even if roles seem to be rather defined, in the end of the day indicators respond to predefined political requests. On the other hand, it is interesting to note how this workshop worked as a space for clarifying positions and what the relevant expertise is.

Workshops are interesting in showing the controversies, and even if that is not the case for all indicators, the excellence one has gone through a certain level of peer review, revision and criticism. Even when you want to have an open underpinning, as a commissioning policy body, you’re in a difficult position: how do you select experts? User 2

Although the aim was reviewing and validating, **people came up with another set of variables [different from the one proposed by the EG] that should have been taken into consideration. People make a difference and that is clear.** Developer 3

Hence, these quotes seem to suggest that indicators are based on selected ‘facts’ of the selected ‘experts’ that are called upon to perform the exercise. The call for evidence-based policy needs to acknowledge this context and carefully examine ‘factual’ promises that cannot be accomplished, which put unnecessary pressures on developers, as well:

You have to understand, we had to consider the request They [DG RTD] just wanted a ranking of member states in terms of this kind of excellence concept. This is what they want; this is what we had to deliver within the project. Developer 1

We found two elements intrinsic to negotiation processes behind indicators development: first, different actors (developers vs. policy-makers) move in different arenas (academic vs. political) and are moved by different interests; second, power relationships set what needs to be measured which make indicators not much more than mere political devices, coherent with a performative function.

4.3 Reimagining what?

Our interviewees explored practical strategies to deal with the policy need for research quality assessments. As researchers, we had assumed that because of many controversies and expressed discontent, there would be a lot of ideas about novel ways to look into the quality of research. Yet, our empirical material shows that there are no clear alternative proposals to either measuring ‘excellent research’ or to enhance the robustness of indicators, except for small variations. As frequently emerged throughout almost all the interviews, many actors highlighted the necessity of carefully interrogating the very use of excellence as the right proxy to research quality, as in this quote:

The **norm of quality** in research that you consider valid and others might not consider valid needs to be discussed as well. A debate is possible and is fundamental within indicators. Developer 2

Despite different positions about the controversial underpinnings of research excellence, widely discussed by the majority of interviewees from each of the three categories, none offered slight or indirect suggestions on how to go beyond the issue of quantification of research quality for policy purposes:

When you have evidence based policy, unfortunately, at the moment, almost the only thing that counts is quantitative data. Case studies and evaluation studies are only strong if they have quantitative data. Then you will get into indicators and it is very difficult to get away from it. User 1

This observation summarises an inevitable commitment to quantification: when asked about research excellence, different actors tend to digress around specific implementations and their implications but do not question in a strong manner the overall scope of the indicator as a means to map or ascertain scientific quality. But, quantifications fit the policy (and political) purpose they are meant to support, as suggested in this honest account by one user:

I think the reasoning is very simple. If we want an indicator that fits its purpose, **which are political purposes**, for policy makers and objective measures, we need to be very clear on what we measure and, as you say, to have the best matching and mismatching between purpose and reality. I think that is the first

question. Then we have to deal with the nitty gritty and see how, sorry, nitty gritty is important, whether we can improve statistically what we have. User 2

Hence, in our interviews the narrative of ‘need for quantification’ inevitability persisted despite the recognition of its inherent limitations and misrepresentations. Interviewees focused on the possibility of improving indicators’ resonance with quality research, avoiding oversimplifications and limiting possible unwanted implications. This quote suggests that the limits of known imperfections of indicators can actually help with raising questions, and therefore we suggest that indicators could be viewed as the prompts to enquire further and not answering devices:

The point is that to take into account the fact that an indicator will never satisfy the totality of the issues concerned, my view is that an indicator is fine when it is built carefully, as long as it is then used not only to provide answers but to raise questions. [...] for example, the indicator we are talking about is fine only as long as it goes along with the discussion of what it does not cover, of what it may hide or not consider with sufficient attention; or in what sense other types of institution or objectives can be taken into account. Developer 2

Along these lines, the importance of allowing for frequently (re)adjustments of evaluation exercises and practices that sustain research indicators is seen as a major improvement:

I am more interested in making sure that as someone involved in composite indicator development, I get the chance to revisit regularly an index which was developed. I can look around and have some kind of conceptual, methodological or statistical review, or see if it is reflecting the ongoing discussions. I can do this if I have the freedom to choose my research. This is not necessarily the case in settings where research is very politically or policy driven. Developer 1

The issue of data availability is quite relevant, not only because of the quality of the built indicators, but more interestingly because existing datasets determine what can be measured and ultimately give shape to the indicator itself, which is a normative issue *tout court*:

Many researchers or many users criticize existing indicators and say they are too conservative. [While they are] always criticized, it is difficult to come with new metrics and the traditional ones are very well grounded in statistics. We have a very good database on data metrics and patents, therefore these databases have some gravitational attraction, and people always go back to them. An indicator needs to be based on existing data. These data has to be acknowledged and there needs to be some experience of them and a bit of time lag between the coverage of new developments by data and then the use for developing indicators. User 4

Finally, excellence does not necessarily need to be a comparative concept, and indeed comparisons ultimately rely on a fair amount of de-contextualisation, which imply overlooking scientific (foremost disciplinary) differences of epistemic, ontological, and practical nature. This is recognised by many of our interviewees:

[Excellence] it is not so useful for comparing EU performance to non-European countries, to US and Japan, because they do not have the same components. They do not have ERC grants, for example! User 4

My suspicion is that [excellence] also depends on the discipline! Practitioner 2

Our quest for reimagination stayed mostly confined to discussing the processes of indicators development, with interviewees largely sharing stances on the apparent inevitability of quantification of research excellence for policy purposes. In fact, we are somehow disappointed that the discussion on other ways to describe and map quality in science did not really produce substantial alternatives. However, few points were raised as central to strengthen the robustness of existing indicators: first, evaluation exercises that deploy research indicators should be frequently checked upon and fine-tuned if necessary; second, what is possible to evaluate should not be constrained by existing datasets but other sources of information should be sought, created, and imagined. In any case, available sources of information are not sufficient when one considers the changing nature of current knowledge production and governance modes which today involve a wider range of societal actors and practices (e.g. knowledge production systems appearing outside mainstream institutions).

5. Discussion

In this article, we explored the making of a particular ‘research excellence’ indicator, starting from its underlying concept and institutional framing. Below, we summarise our findings through five main points. Together, these may constitute departing points for future debates around alternative evaluation framings, descriptors, and frameworks to describe and map the quality of scientific research in the EU.

5.1 Research excellence: contested concept or misnomer?

Early in this article, we advanced the idea of excellence as an essentially contested concept, as articulated by Gallie (1955). Our interviews seem to concur with the general idea that the definition of such a concept does not seem to be stable and that there are genuine difficulties (and actual unpreparedness) among interviewees even to come up with a working definition for ‘research excellence’. In most cases, interviewees seem to agree that research excellence is a multi-dimensional, complex, and value-laden concept whose quantification is likely to end in controversy. ‘Institutionalised’ definitions, which may not necessarily be subject of a thorough reflection, were often given by our interviewees; they repeatedly remarked that each definition depends very much on the actors involved in developing indicators. So, would more extended debate about the meanings and usefulness of the concept to assess and compare scientific research quality be helpful to address some of the current discussions?

5.2 Inescapability of quantification?

The majority of our interviewees had a hard time imagining the assessment of research that is not reliant on *quantification*. Yet, quantifying or not quantifying research excellence for policy purposes does not seem to be the question, the issue rather revolved around *what* really needs to be quantified. Is the number of published papers really an indication of excellence? Does paper citation really imply its actual reading? As with classifications (Bowker and Star 1999), indicators of research excellence are both hard to live with and without. The question is how to make life with indicators acceptable while recognising their fallibility. Recognising that

quantifying research excellence requires choices to be made, then the values and interests of such choices serve at the neglect of others becomes an important reflection. We would argue that quantifying research excellence is first and foremost a political and normative issue and as such, Merton's (1973) pertinent question 'who is to judge on research excellence?' remains.

The need for quantification is encouraged by and responds to the trend of evidence-based policy. After all, this is a legacy of the 'modern' paradigm for policy making which needs to be based on scientific *evidence* and this, in turn, needs to be delivered in numbers. However, as Boden and Epstein, (2006) remarked, we might be in a situation of 'policy based evidence' instead, where scientific research is assessed and governed to meet policy imaginaries of scientific activity (e.g. focus on outcomes such as the number of publications, 'one size fits all' approaches to quantification across different scientific fields, etc.). The question then remains, that is, can ideas of qualifying quantifications be developed also in this case?

5.3 The lamppost

In Mulla Nasraddin's story, the drunken man tries to find under the lamppost the keys he lost. Some of the interviewees suggested that the bottleneck for quantification is existing data. In other words, data availability influences what is possible to quantify: only those parameters for which there are already considerable data, that is those which are easy to count seem to be the ones taken into account. We argue that this type of *a priori* limitation needs to be reflected upon, not least because knowledge production and the ways in which researchers make visible their work to the public are not confined to academic formats only. Moreover, if one considers the processes by which scientific endeavour actually develops, then we might really need to see outside the lamppost's light circle. Can we afford to describe and assess 'excellent research' exclusively relying on current parameters for which data are already available?

5.4 Drawing hands

In an introductory piece about the co-production idiom, Jasanoff (2004: 2) says that 'the ways in which we know and represent the world are inseparable from the ways we choose to live in it'. We concur with the idea that the construction of indicators is a sociopolitical practice. From such a perspective, it becomes clear that knowledge production practices are in turn conditioned by knowledge production *assessment* practices, exactly as depicted in artist M. C. Escher's piece *Drawing Hands*. In other words, whichever ways (research excellence) indicators are constructed, their normative nature contributes to redefining scientific practices. We suggest that the construction of an indicator is a process in which both the concept (research excellence) and its measurements are mutually defined and are co-produced. If an indicator redefines reputation and eligibility for funding, researchers will necessarily adapt their conduct to meet such pre-established standards. However, this understanding is not shared by all interviewees, which suggests that future practice needs to raise awareness of the normativity inherent to the use of indicators.

5.5 One size does not fit all

Indicators necessarily de-contextualize information. Many of our interviewees suggested that other types of information would need to be captured by research indicators; to us this casts doubts about the appropriateness of the use of indicators alone as the relevant devices for assessing research with the purposes of designing policy.

What do such indicators tell us about scientific practices across fields and different countries and institutions? The assumption that citation and publication practices are homogenous within different specialties and fields of science has been previously demonstrated as problematic (Leydesdorff 2008), and it is specifically within the policy context that indicators need to be discussed (see e.g. Moed et al. 2004).

The STS literature offers us examples of cultures of scientific practice that warn us that indicators alone cannot be used to sustain policies, but they certainly are very useful to ask questions.

Nowotny (2007) and Paradeise and Thoenig (2013) argued that like many other *economic* indicators, 'research excellence' is promoted at the EU level as a 'soft' policy tool (i.e. it responds to benchmarks to compel Member States to meet agreed obligations). But the implied measurements and comparisons 'at all costs' cannot be considered 'soft' at all: they inevitably trigger unforeseen and indirect incentives in pursuing a specific kind of excellence (see e.g. Martin 2011) often based on varied, synthetic and implicit evaluations. In the interviews, stories were told to us, about purposeful retuning of indicators because some countries would not perform as expected when variations to the original indicators were introduced.

If going beyond quantification eventually turns out not being an option at all, at least we should aim for more transparency in the 'participatory' processes behind the construction of indicators. To cite Innes' words 'the most influential, valid, and reliable social indicators are constructed not just through the efforts of technicians, but also through the vision and understanding of the other participants in the policy process. Influential indicators reflect socially shared meanings and policy purposes, as well as respected technical methodology' (Innes 1990).

6. Final remarks

This work departed from the idea that the concept of research excellence is hard to be institutionalised in the form of stable research excellence indicators, because it inevitably involves endless disputes about its usage. Therefore, we expected to find alternative by other imaginaries and transformative ideas that could sustain potential changes. To test these ideas, we examined the RES&T indicator development and its quantification, highlighting that this indicator is developed in a context in which it simultaneously respond to and normalise both scientific practice and policy expectations. We also explored the difficulties of measuring a concept (research excellence) that lacks agreed meanings. The in-depth interviews conducted with relevant actors involved in the development of the RES&T research indicator suggest that, while respondents widely acknowledge intrinsic controversies in the concept and measurement, and are willing to discuss alternatives (what we called 're-imagination'), they did not find it easy to imagine alternatives to address research quality for policy purposes. Quantification is hard-wired into practices and tools to assess and assure the quality of scientific research that are further reinforced by the current blind and at-all-costs call for quantified evidence based policy to be applied in twenty-eight different EU Member States. However, suggestions were made to make reimagination a continuous stage of the process of developing excellence assessments, which reminds us of Barré's agora model (Barré 2004).

To conclude, more than a contested concept, our research lead us to wonder whether 'research excellence' could be a misnomer to assess the quality of scientific research in a world where processes,

and not only outcomes, are increasingly subject of ethical and societal scrutiny? Or, what is the significance of excellence indicators when scientific research is a distributed endeavour that involves different actors and institutions often even outside mainstream circles?

Conflict of interest statement. The views expressed in the article are purely those of the writers and may not in any circumstances be regarded as stating an official position of the European Commission and Robobank.

Notes

1. The official Horizon2020 document defines that research excellence is about to “[...] ensure a steady stream of world-class research to secure Europe’s long-term competitiveness. It will support the best ideas, develop talent within Europe, provide researchers with access to priority research infrastructure, and make Europe an attractive location for the world’s best researchers.” (European Commission, 2011) (p.4).
2. From “Against Excellence”, the Guardian, 19 /12/2014 Retrieved at <https://www.theguardian.com/science/political-science/2014/dec/19/against-excellence>.

Acknowledgements

The authors would like to thank the contribution of the interviewees, the comments and suggestions of two anonymous reviewers as well as the participants of the workshop on “Excellence Policies in Science” held in Leiden, 2016. The views expressed in the article are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

References

- Barré, R. (2001) ‘Sense and Nonsense of S&T Productivity Indicators’, *Science and Public Policy*, 28/4: 259–66.
- Barré, R., (2004) ‘S&T indicators for policy making in a changing science-society relationship’, in: H. F., Moed, W., Glänzel, and U. Schmoch. (eds) *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, pp. 115–131. Dordrecht: Springer.
- Barré, R. (2010) ‘Towards Socially Robust S&T Indicators: Indicators as Debatable Devices, Enabling Collective Learning’, *Research Evaluation*, 19/3: 227–31.
- , Hollanders, H., and Salter, A. (2011). *Indicators of Research Excellence*. Expert Group on the measurement of innovation.
- Benett, K. (2016) ‘Universities are Becoming Like Mechanical Nightingales’. *Times Higher education*. <<https://www.timeshighereducation.com/blog/universities-are-becoming-mechanical-nightingales>>
- Boden, R., and Epstein, D. (2006) ‘Managing the Research Imagination? Globalisation and Research in Higher Education’, *Globalisation, Societies and Education*, 4/2: 223–36.
- Boulanger, P.-M. (2014) *Elements for a Comprehensive Assessment of Public Indicators*. JRC Scientific and Policy Reports, Luxembourg: Publications Office of the European Union.
- Bowker, G. C., and Star, S. L. (1999) *Sorting Things Out: Classification and its Consequences*. Cambridge: MIT Press.
- Butler, L. (2003a) ‘Explaining Australia’s Increased Share of ISI Publications—the Effects of a Funding Formula Based On Publication Counts’, *Research Policy*, 32: 143–55.
- Charmaz, K. (2006) *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*, Vol. 10. <http://doi.org/10.1016/j.lisr.2007.11.003>
- Clarke, A. E. (2003) ‘Situational Analyses’, *Symbolic Interaction*, 26/4: 553–76.
- Collier, D., Daniel Hidalgo, F., and Olivia Maciuceanu, A. (2006) ‘Essentially Contested Concepts: Debates and Applications’, *Journal of Political Ideologies*, 11: 211–46.
- Desrosières, A., (2015) ‘Retroaction: How Indicators Feed Back onto Quantified Actors’. In: Rottenburg, et al. (eds) *The World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge: Cambridge University Press.
- Elzinga, A. (2012) ‘Features of the Current Science Policy Regime: Viewed in Historical Perspective’, *Science and Public Policy*, 39/4: 416–28.
- European Commission (2014) *Innovation Union Competitiveness Report 2013—Commission Staff Working Document, Directorate-General for Research and Innovation*. Luxembourg: Publications Office of the European Union.
- Espeland, W. (2015) ‘Narrating Numbers’. In: Rottenburg et al. (eds) *The World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge, UK: Cambridge University Press.
- Funtowicz, S. O., and Ravetz, J. R. (1990) *Uncertainty and Quality in Science for Policy*, 229–41. Dordrecht: Kluwer Academic Publishers.
- Gallie, W. B. (1955) ‘Essentially Contested Concepts’, *Proceedings of the Aristotelian Society*, 56: 167–98.
- Gieryn, T. F. (1983) ‘Boundary-Work and the Demarcation of Science from Non-science: Strains and Interests in Professional Ideologies of Scientists’, *American Sociological Review*, 48/6: 781–95.
- Grupp, H., and Moge, M. (2014) ‘Indicators for National Science and Technology Policy. How Robust are Composite Indicators?’, *Research Policy*, 33/2004: 1373–84.
- , and Schubert, T. (2010) ‘Review and New Evidence on Composite Innovation Indicators for Evaluating National Performance’, *Research Policy*, 39/1: 67–78.
- Gunasekara, C. (2007) ‘Pivoting the Centre: Reflections on Undertaking Qualitative Interviewing in Academia’, *Qualitative Research*, 7: 461–75.
- Hardeman, S. (2013) ‘Organization Level Research in Scientometrics: A Plea for an Explicit Pragmatic Approach’, *Scientometrics*, 94/3: 1175–94.
- , Van Roy, V., and Vertesy, D. (2013) *An Analysis of National Research Systems (I): A Composite Indicator for Scientific and Technological Research Excellence*. JRC Scientific and Policy Reports, Luxembourg: Publications Office of the European Union.
- Hessels, L. K., and van Lente, H. (2008) ‘Re-thinking New Knowledge Production: A Literature Review and a Research Agenda’, *Research Policy*, 37/4: 740–60.
- Hicks, D. (2012) ‘Performance-based University Research Funding Systems’, *Research Policy*, 41/2: 251–61.
- Innes, J. E. (1990) *Knowledge and Public Policy. The Search for Meaningful Indicators*. New Brunswick (USA) and London (UK): Transaction Publishers.
- Jasanoff, S. (ed.) (2004) *States of Knowledge: The Co-Production of Science and the Social Order*. London: Routledge.
- Lamont, M. (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge/London: Harvard University Press.
- Leydesdorff, L. (2008) ‘Caveats for the Use of Citation Indicators in Research and Journal Evaluations’, *Journal of the American Society for Information Science and Technology*, 59/2: 278–87.
- Martin, B. R. (2011) ‘The Research Excellence Framework and the ‘impact agenda’: are we Creating a Frankenstein Monster?’, *Research Evaluation*, 20/3: 247–54.
- (2013) ‘Whither Research Integrity? Plagiarism, Self-Plagiarism and Coercive Citation in an Age of Research Assessment’, *Research Policy*, 42/5: 1005–14.
- Merton, R. K. (1973) ‘Recognition and Excellence: Instructive Ambiguities’. In: R. K., Merton (ed.) *The Sociology of Science*. Chicago: University of Chicago Press.
- Mood, H.F., Glänzel, W., Schmoch, U. (eds) (2004) *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*. Dordrecht: Kluwer.
- Nowotny, H. (2007) ‘How Many Policy Rooms Are There? Evidence-Based and Other Kinds of Science Policies’, *Science, Technology & Human Values*, 32/4: 479–90.

- Paradeise, C., and Thoenig, J.-C. (2013) 'Organization Studies Orders and Global Standards', *Organization Studies*, 34/2: 189–218.
- Polanyi, M. (2000) 'The Republic of Science: Its Political and Economic Theory', *Minerva*, 38/1: 1–21.
- Porter, T. M., (2015) The Flight of the Indicator. In: Rottenburg et al. (eds) *The World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge: Cambridge University Press.
- Rafols, I., Leydesdorff, L., O'Hare, A. et al. (2012) 'How Journal Rankings Can Suppress Interdisciplinary Research: A Comparison Between Innovation Studies and Business & management', *Research Policy*, 41 /7: 1262–82.
- Saltelli, A., D'Hombres, B., Jesinghaus, J. et al. (2011) 'Indicators for European Union Policies. Business as Usual?', *Social Indicators Research*, 102/2: 197–207.
- , and Funtowicz, S. (2017). 'What is Science's Crisis Really About?' *Futures*—in press. <<http://www.sciencedirect.com/science/article/pii/S0016328717301969>> accessed July 2017.
- Sarewitz, D. (2015) 'Reproducibility Will Not Cure What Ails Science', *Nature*, 525: 159.
- Smith, A. F. (1996) 'MAD Cows and Ecstasy: Chance and Choice in an Evidence-Based Society', *Journal-Royal Statistical Society Series A*, 159: 367–84.
- Sørensen, M. P., Bloch, C., and Young, M. (2015) 'Excellence in the Knowledge-Based Economy: from Scientific to Research Excellence', *European Journal of Higher Education*, 1–21.
- Stilgoe, J. (2014). 'Against Excellence'. *The Guardian*, 19 December 2014.
- Tijssen, R. J. (2003) 'Scoreboards of Research Excellence', *Research Evaluation*, 12/2: 91–103.
- Vertesy, D., and Tarantola, S. (2012). *Composite Indicators of Research Excellence*. JRC Scientific and Policy Reports. Luxembourg: Publications Office of the European Union.
- Weinberg, A. M. (2000) 'Criteria for Scientific Choice', *Minerva*, 38/3: 253–66.
- Weingart, P. (2005) 'Impact of Bibliometrics upon the Science System: Inadvertent Consequences?', *Scientometrics*, 62/1: 117–31.
- Wilsdon, J. (2015) 'We Need a Measured Approach to Metrics', *Nature*, 523/7559: 129.