Open-Access Repositories Worldwide, 2005–2012: Past Growth, Current Characteristics, and Future Possibilities

Stephen Pinfield, Jennifer Salter, and Peter A. Bath
Information School, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP, United Kingdom. E-mail: {S.Pinfield, P.A.Bath}@Sheffield.ac.uk, JASalter1@sheffield.ac.uk

Bill Hubbard, Peter Millington, Jane H.S. Anders, and Azhar Hussain
Centre for Research Communications, University of Nottingham, Greenfield Medical Library, Nottingham, NG7 2UH, United Kingdom. E-mail: {Bill.Hubbard, Peter.Millington, Jane.Anders, Azhar.Hussain}@Nottingham.ac.uk

This paper reviews the worldwide growth of open-access (OA) repositories, 2005 to 2012, using data collected by the OpenDOAR project. Initial repository development was focused on North America, Western Europe, and Australasia, particularly the United States, United Kingdom, Germany, and Australia, followed by Japan. Since 2010, there has been repository growth in East Asia, South America, and Eastern Europe, especially in Taiwan, Brazil, and Poland. During the period, some countries, including France, Italy, and Spain, have maintained steady growth, whereas other countries, notably China and Russia, have experienced limited growth. Globally, repositories are predominantly institutional, multidisciplinary and English-language based. They typically use open-source OAI-compliant software but have immature licensing arrangements. Although the size of repositories is difficult to assess accurately, available data indicate that a small number of large repositories and a large number of small repositories make up the repository landscape. These trends are analyzed using innovation diffusion theory, which is shown to provide a useful explanatory framework for repository adoption at global, national, organizational, and individual levels. Major factors affecting both the initial development of repositories and their take-up include IT infrastructure, cultural factors, policy initiatives, awareness-raising activity, and usage mandates. Mandates are likely to be crucial in determining future repository development.

Received August 16, 2013; revised September 13, 2013; accepted September 15, 2013

 • Published online 28 April 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23131

Introduction

Open-access (OA) repositories of academic and research outputs are a significant development in the 21st-century global research information environment. This paper reviews the worldwide growth of OA repositories between December 2005 and December 2012. It also provides an overview of the main characteristics of the global repository infrastructure at the end of that period. It does so based on data collected during the 7-year period by the *Open*DOAR project based in the Centre for Research Communications at the University of Nottingham, United Kingdom. In addition to outlining historic trends and current characteristics of the repository landscape, the paper also locates these developments within an overarching explanatory framework and discusses factors affecting possible future trajectories of repository development.

Definitions

It is important initially to establish some definitions of key concepts, particularly *open access* and *repositories*, because these form the basis of criteria for inclusion (and exclusion) in this study. Open access may be defined as where digital content is fully, freely, immediately, and permanently available and can be viewed and reused with minimal restrictions (adapted from Pinfield, 2009, p. 166; with the explication that follows also adapted from the same source). The *digital content* referred to in the definition may take a number of different forms, including text, data, and rich media (such as video). This content is *fully available*, meaning that the content is accessible in its entirety rather than in a partial form (in the case of a journal article, for example, this would mean the full text is available and not

just the abstract). It is also *freely available*, meaning there are no costs charged at the point of access (although costs are of course incurred in generating and disseminating the content). The content is *immediately available*, meaning there are no formal delays or embargoes on its availability. It is also *permanently available*, meaning it is available on an ongoing basis so that it can be accessed and cited (usually involving a commitment to maintaining persistent access paths to as well as preservation of the content itself). The content may be *viewed* and also *reused with minimal restrictions*. *Ability to view* the content means it can be read (normally by humans), and *ability to reuse* the content means it can be copied, mined, processed, analyzed, indexed, and redistributed (often by machines) with minimal rights restrictions.

The ideas characterized in this definition as viewing and reusing content are at the center of Suber's (2008, 2012) important distinction between different *subspecies* of OA. After providing a brief definition of OA as "digital, online, free of charge, and free of most copyright and licensing restrictions" (Suber, 2012, p. 4), Suber goes on to distinguish between, on one hand, "Gratis open access" (which is free to view), and on the other hand, "Libre open access" (free to reuse and rework as well as view):

> "Gratis OA is free of charge but not freer than that. Users must still seek permission to exceed fair use. Gratis OA removes price barriers but not permission barriers.
>
> Libre OA is free of charge and also free of some copyright and licensing restrictions. Users have permission to exceed fair use, at least in certain ways . . . Libre OA removes price barriers and at least some permission barriers." (Suber, 2012, p. 66)

The repositories recorded in *Open*DOAR and included in this study comply in general terms with the definitions of OA provided by Pinfield and Suber, including both Gratis and Libre OA, although formerly-embargoed content is also permitted (see below) if it is made available on an OA repository.

The definition of *repository* used here follows Pinfield (2009):

> "A repository may be defined as a set of systems and services that facilitates the ingest, storage, management, retrieval, display, and reuse of digital objects. Repositories may be set up by institutions, subject communities, research funders, or other groups. They may provide access to a variety of digital objects, including peer-reviewed journal articles, book chapters, theses, datasets, learning objects, or rich media files." (Pinfield, 2009, p. 165)

The key distinction often drawn in the literature in relation to academic and research repositories is between disciplinary repositories and institutional repositories (IRs) (Nicholas, Rowlands, Watkinson, Brown, & Jamali, 2012; Xia, 2008). The former contain the outputs of a particular subject community; well-known examples include arXiv for high-energy physics, RePEc for economics, and PubMed Central for biomedical and life sciences. The latter house the outputs of researchers and other research materials from a particular institution; early examples include Caltech in the United States and the University of Southampton in the United Kingdom. To these, Nicholas et al. (2012) also add "format" repositories that contain particular types of files, such as e-theses, data sets, or learning objects. There are also a small number of *governmental* repositories managed by national governments and government-sponsored agencies, and *aggregating* repositories that present content harvested from other sources. All of these categories of repositories are recorded in *Open*DOAR: disciplinary, institutional, aggregating, and governmental as *repository types*, with formats being recorded as different *content types*.

Research Context

Academic and research OA repositories have existed in various forms for more than two decades and have been featured in the research and practitioner literature since then. As well as formally published works, discussion around policy and practice in this area has taken place in e-mail discussion lists, blogs, and other more informal venues. Increasingly, the latest thinking (at least in provisional form) first appears in those places (as is increasingly the case in a variety of fields), and so it is important to take them into account in examining the research context.

Much of the early discussion centered on now well-established, disciplinary repositories such as arXiv, setup in August 1991 (e.g., Ginsparg, 1997; Halpern & Lagoze, 1999; McKiernan, 2000; Pinfield, 2001). However, there has been a rapid growth of repositories more recently enabled to a significant degree by two technical developments between 2000 and 2003 that were widely discussed in the literature and other fora. The first of these was the development of the Open Archives Initiative Protocol for Metadata Harvesting (OAI PMH), an interoperability protocol for exchanging information between repositories (Lagoze & Sompel, 2003; Sompel & Lagoze, 2000). The second development (arising from the first) was the release of OAI-PMH-compliant open-source repository software, such as ePrints (Gutteridge, 2002) and DSpace (Smith et al., 2003). These developments meant that installing a repository became technically straightforward and therefore facilitated the growth of repository numbers, including the development of institutional repositories (Crow, 2002; Lynch, 2003). Repository growth began to accelerate in the middle of the first decade of the 21st century on the foundation of these two developments, and *Open*DOAR (Directory of Open Access Repositories) was set up to track that growth.

During this period, there was also a maturing of new models of scholarly communication particularly associated with OA. Many OA repositories began to be designed to support so-called Green OA: the depositing in repositories of research papers that have been or are expected to be published in peer-review journals, often involving authors self-archiving their own outputs (Björk, Laakso, Welling, &

DOI: 10.1002/asi

Paetau, 2013; Suber, 2012). There is now a large literature on OA repositories, most of it focusing in particular on their role in supporting Green OA as a form of scholarly communication. Studies cover a wide range of issues including user attitudes and behaviors (including Kim, 2010), different disciplinary positions (such as Xia, 2007), changing roles for librarians (including Walters, 2007), and legal issues (e.g., Hanlon & Ramirez, 2011). There are also various practice-based case studies (such as Connell, 2011) and guides, particularly on institutional repositories (e.g., Nabe, 2009). Some research has analyzed growth trends and key characteristics of repositories often using *Open*DOAR data. A number of studies have focused on IRs in specific geographical areas, such as Canada (Shearer, 2006), Zimbabwe (Nyambi & Maynard, 2012), India (Sawant, 2011; Tripathi & Jeevan, 2011), or Asian countries in general (Nazim & Mukherjee, 2011). Other focuses have been on subject-based repositories for particular disciplines, such as chemistry (Warr, 2003) and computer science and IT (Bhat, 2010). There is also a significant amount of literature on format repositories, particularly repositories for theses (an overview has been provided by Yiotis, 2008) and learning objects (a quantitative study of the global infrastructure is provided by Ochoa and Duval, 2009). All of these areas of the literature contribute to the discussion on the (potential) role of OA repositories in a developing scholarly communication infrastructure.

In the discussion, Green OA is often set against *Gold* OA, open-access publication in journals. For example, Harnad, a well-known Green advocate, following the new emphasis in U.K. policy on Gold OA in 2012, referred rhetorically in his blog and e-mail discussion lists to "fools Gold," criticizing what he saw as a "Gold Rush," and calling for a greater emphasis on Green OA (e.g., Harnad, n.d.). However, as Pinfield (2009) notes, repositories and journals are not necessarily mutually exclusive OA solutions. For example, some large-scale repositories, notably PubMed Central, have been set up primarily to house material first published in an OA form by journals. Such repositories are covered by the *Open*DOAR database along with others.

Running through much of the literature and other fora is an awareness that even at a time of growth of repository numbers there has been relatively slow take-up of repositories by researchers (in terms of contributing and using content) in most areas (with repositories such as arXiv and PubMed Central being exceptions). A large amount of the practitioner output on repositories (particularly IRs) focuses on this issue (e.g., Davis & Connolly, 2007; Lawal, 2002) and includes case studies of awareness-raising and advocacy programs aimed at encouraging researcher engagement with repositories (e.g., Buehler, 2013; Foster & Gibbons, 2005; Palmer, Teffeau, & Newton, 2008). Related to this, there has also been extensive discussion on the development of policy mandates by institutions, funders, or other agencies as a means of increasing usage (e.g., Kennan, 2011; Larivière, Gingras, Carr, Brody, & Harnad, 2012; Pinfield, 2005).

The research literature includes various analyses of repository development deploying different theoretical models particularly focusing on the issue of the usage levels of individual researchers. Kim (2010, 2011) uses socio-technical interaction networks (STIN) and social exchange theory to study faculty self-archiving practices. Using a combination of qualitative and qualitative data from surveys and interviews, she identifies a number of factors that have implications for repository implementation. She concludes that "Faculty are motivated by OA advantages to users, disciplinary norms, and no negative influence on academic reward. However, barriers to self archiving—concerns about copyright, extra time and effort, technical ability, and age—imply the provision of services to assist faculty with copyright management, and with technical and logistical issues, could encourage higher rates of self-archiving" (Kim, 2010, p. 1909).

Another theoretical framework usefully deployed to explain repository adoption by individual researchers is the unified theory of acceptance and use of technology (UTAUT). First developed by Venkatesh, Morris, Davis, and Davis (2003), this theory has been used by Mann, von Walter, Hess, and Wigand (2009) and Dulle and Minishi-Majanja (2011) whose quantitative studies demonstrate the particular importance of "Performance Expectancy," "the degree to which an individual believes that using the system will help him or her attain gains in job performance" (Venkatesh et al., 2003, p. 447). Other important factors, including attitude to OA, awareness of OA developments, and expected effort of using the technology are also identified as significant by Dulle and Minishi-Majanja (2011). However, as Björk et al. (2013) observe, the major weakness of these studies is that they conflate Green and Gold OA in their analyses. A more recent study by Singeh, Abrizah, and Karim (2013) casts some doubt on the utility of UTAUT in this context because the authors do not find a correlation between UTAUT determinants and use behavior associated with repositories. However, they still recommend further research to test their conclusions, because, as they acknowledge, theirs is a small-scale study and it does at least find a positive relationship between the determinants. All of the studies make practical recommendations around improving researcher awareness and considering the potential of OA policy mandates.

A final theoretical model used to explain repository development is innovation diffusion theory (IDT). This has been deployed by Jones, Andrew, and MacColl (2006) in a practitioner context to frame a discussion of the development of an institutional-level advocacy campaign aimed at promoting the use of an institutional repository. IDT has also been used at a different level by Xia (2012) to explain different adoption patterns of both OA journals and repositories worldwide. His analysis usefully incorporates discussion of technological, cultural, and policy factors that influence varying take-up globally. These studies using IDT illustrate a particular strength of diffusion theory in that it can be used to explain adoption characteristics at a number

DOI: 10.1002/asi

of levels, including the local level (taking into account the behaviors of individual actors) and the global level (taking into account large-scale cultural and technological trends). It is deployed later in this paper to provide an explanatory framework for the data presented, which relate to the levels already described but also, crucially, to organizational-level adoption.

This article provides a high-level analysis of the entire OA repository population worldwide, including all repository types (disciplinary, institutional, format, and others). It reviews the major growth trends of repositories across different continents and countries and also analyzes repositories' current characteristics. It makes a number of observations about the contribution to and usage of repositories within research communities. Based on the data presented, it discusses various factors that have affected repository development to date and are likely to impact on further growth in the future.

Data Collection

The data analyzed here were gathered as part of the *Open*DOAR (Directory of Open Access Repositories) project based at the University of Nottingham, United Kingdom, carried out initially in partnership with Lund University, Sweden. *Open*DOAR was set up in 2005 to create "an authoritative directory of academic open access repositories" throughout the world (OpenDOAR, n.d.-a). It began collecting data in that year, which were used as the basis of the *Open*DOAR web directory, launched in December 2005 and available since then. The records are periodically reviewed, updated and augmented in the database and therefore the online service provides an up-to-date snapshot of the worldwide academic repository landscape.

The *Open*DOAR service is now used as an important source for data research on OA developments (e.g., Björk et al., 2013), OA advocacy (e.g., Morrison, 2012), and also via its application programming interface (API) as a data feed for other OA-related online services (including BASE, CORE, Institutional Repository Search, Open Access Tracker, OpenAIRE, OpenDepot, Repository 66, and ROAR). Because of its widespread use, it is useful to provide here some detail on its design and to review the repository growth trends and characteristics it records.

An *Open*DOAR online public record consists of the following:

- Organization: the official title and country of origin of the organization(s) responsible for the repository
- Description: a free text description of the repository and its services
- OAI PMH: the Base URL for the Open Archives Initiative Protocol for Metadata Harvesting
- Software: the software used as the basis for the repository, when known
- Size: the number of records in the repository (updated every 2 to 4 weeks)
- Subjects: broad subject description (using a controlled vocabulary based on the broad categories in the Library of Congress classification scheme) or *multidisciplinary* if covering a wide range of subjects
- Content: content types based on controlled vocabulary descriptions of "articles," "conferences," "books," "theses," "learning objects," and other major formats
- Languages: principal languages present in the repositories content
- Policies: descriptions of submission and reuse rights for metadata and full items, and content on preservation policies where available, using standardized vocabulary
- Remarks: free text field providing any other additional information, for example on partnerships or membership of consortia
- *Open*DOAR ID: the unique record number within the *Open*DOAR database

These key data points are the basis of this study. Added to them are other data recorded in *Open*DOAR, including interoperability protocols deployed by repositories and geographical coordinates of systems.

*Open*DOAR records are, for the most part, created and maintained manually, rather than by automated harvesting, to ensure a high level of quality. The *Open*DOAR website explains this approach:

> "As part of this quality assurance, *Open*DOAR administrators carry out a detailed investigation of each site, to ensure that the site is both accessible and provides access to full-text content, without the need for a subscription or registration." (OpenDOAR, n.d.-b)

This manual record creation is augmented by regular automated checks on each repository site recorded in the database to ensure it is still available and also to enable the update of the "Size" field in *Open*DOAR. Repositories, which prove to be inaccessible to the *Open*DOAR robot, are then rechecked manually and any necessary changes to the Directory made, including the removal of records for repositories no longer available. It normally takes between 1 and 2 weeks from initial suggestion or discovery of a repository to its inclusion in the *Open*DOAR database. This has changed little since the first year of *Open*DOAR's release and therefore the database administrators express confidence that *Open*DOAR has been a relatively accurate and up-to-date picture of the repository landscape, certainly since it became fully established in the fourth quarter of 2006.

Potential repositories for inclusion in *Open*DOAR are identified in various ways, including proactive searching by *Open*DOAR staff and checking other sources such as registrations on the OAI website (repositories included in *Open*DOAR do not necessarily have to be OAI-PMH-compliant but most are). In more recent years, many of the new records have been added as a result of managers of recently launched repositories applying to *Open*DOAR to register their service. This community-based participation in *Open*DOAR, also seen in suggested updates to records, has

DOI: 10.1002/asi

undoubtedly played an important role in maintaining the currency of the database. On occasions, some records have been added to *Open*DOAR in batches to update the database rapidly. Although this is relatively rare, it applies to some repositories particularly in non-English-language-speaking countries, such as Japan and Poland. This pragmatic approach to keeping the database as current as possible reflects the fact that it is designed primarily to be an up-to-date directory at any given time rather than an archive of historic trends and it is illustrated in some of the data presented below.

The criteria for inclusion in the database are specified on the *Open*DOAR website, which summarizes its approach as aiming to,

> "collect and provide information solely on sites that wholly embrace the concept of open access to full text resources that are of use to academic researchers. Thus sites where any form of access control prevents immediate access are not included; likewise sites that consist of metadata records only are also declined." (OpenDOAR, n.d.-a)

This pragmatic definition of OA, encompassing both Gratis and Libre OA and also including previously-embargoed content, allows for the inclusion of the widest possible range of OA repositories supporting the research community. It does, however, exclude repositories that require user registration or application for content, conditions that are typical of many data repositories. Repositories must contain a proportion of full-text or other full-content files, but the published criteria do allow for some metadata-only records in repositories:

> "Repositories where both full-text and metadata entries are available in an open access manner, may be included, where all other inclusion criteria are met." (OpenDOAR, n.d.-b)

Thus the records listed in the database in the *Size* field may contain a combination of full-text and other fully open resources, plus metadata-only records. *Open*DOAR does, however, reject a significant proportion of the repositories suggested for inclusion, one reason for which may be a repository entirely populated with metadata-only records:

> "*Open*DOAR currently rejects about one quarter of the suggestions it receives. Rejected sites are primarily spam, metadata only repositories, duplicate suggestions or open access journals." (OpenDOAR, n.d.-b)

The last of these, OA journals, are included in *Open*DOAR's sister service, DOAJ (Directory of Open Access Journals) and are therefore not duplicated in *Open*DOAR. Most repositories listed in *Open*DOAR contain academic papers (e-prints)—the service was originally conceived as registry of e-print repositories. However, a minority of repositories also included in *Open*DOAR contain a wide variety of other formats such as data sets, learning objects, and images.

The approach taken by *Open*DOAR differs from other available directories of OA repositories. For example, ROAR (Registry of Open Access Repositories) is based on automatic harvesting (including records derived from *Open*DOAR itself) rather than manual record creation. One major advantage of the ROAR approach, it might be argued, is immediacy: Regular automated harvesting means there are no delays in record creation. In contrast, the *Open*DOAR team has had to work hard to ensure that its manual record creation is timely in establishing its processes (although it has a good record in doing so). However, a major disadvantage of the entirely automated approach is the fact that it tends to pick up a significant number of invalid sites, including experimental implementations that have few records or those with metadata-only entries. Such sites are rejected by *Open*DOAR. As a result, ROAR tends to have larger numbers of repositories listed in its directory at any one time than *Open*DOAR (a fact observed with partial explanation by Morrison, 2012).

As well as counts of repositories, *Open*DOAR also differs significantly from ROAR in the way it counts items within each repository. Repository sizes in ROAR are determined using the OAI PMH harvesting of record counts. However, testing carried out by *Open*DOAR has found this to be unreliable and it has therefore developed a set of alternative automated counting methods, which testing has shown to be more accurate. A total of 12 different methods have been developed and applied selectively to different repositories depending on a number of factors including, for example, the software that a repository uses (Millington, 2013). For each repository listed in *Open*DOAR, an initial assessment is made of which counting method is likely to yield the most accurate results and, following tests where necessary, this preferred method is then used to harvest data on record counts from that repository. Consequently, *Open*DOAR and ROAR, using different approaches to record counts, usually record different sizes for any given repository.

The data analyzed here were extracted from the *Open*DOAR database and cover the period December 2005 to December 2012. Data on the historic growth of the database during that period were provided by the Centre for Research Communications on January 7, 2013. A snapshot of the entire database was also taken on December 19, 2012, downloaded through the publicly available *Open*DOAR API. In addition, comparable detailed snapshot data from February 15, 2008, were also available, allowing some comparisons to be made. The data were processed and analyzed using SPSS and Excel.

Results

Worldwide Repository Growth

Overall, the total number of repositories in *Open*DOAR grew from 128 in December 2005 to 2,253 in December 2012 (see Figure 1). This represents a 1660% increase

DOI: 10.1002/asi

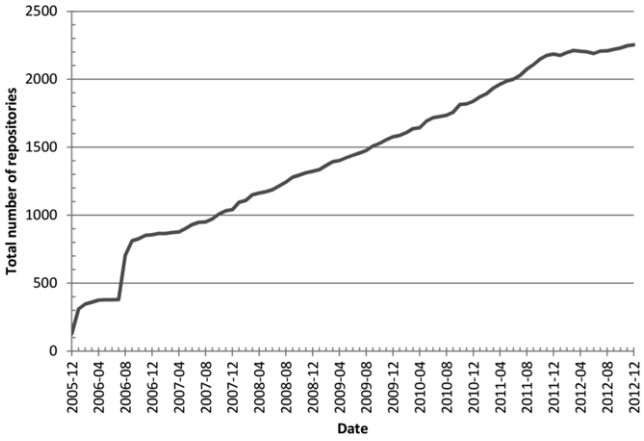

FIG. 1. Overall growth of repositories in *Open*DOAR, December 2005 to December 2012.

during the period. However, the rate of growth was not entirely consistent during that time: There was rapid growth in the first year (from 128 to 855 repositories between December 2005 and December 2006) and a slowing down of growth in the final year (from 2,185 to 2,251 between December 2011 and December 2012).

In relation to the first of these periods, this (2005 to 2006) was one of rapid growth in repository numbers; however, other factors may also be important in explaining the changes. The growth in numbers is to a certain extent undoubtedly a function of *Open*DOAR service development and consolidation during its first year of operation. In particular, the step change noticeable between July and August 2006 (from 379 repositories to 704), preceded by a 3-month plateau, is certainly an artifact of the technical development of the service during that time. Between May and July 2006, *Open*DOAR underwent a technical infrastructure upgrade as part of its move from being a pilot to becoming a full production service. During those 3 months, records for repositories were created off-line in a separate database and batch uploaded into the new upgraded service during August and September 2006. This accounts for much of the rapid rise in record numbers during that period.

However, the apparent slowing down of growth in the final year is more difficult to account for. It is likely to be attributable to a complex set of factors, some to do with repository growth itself, others to do with the management of the *Open*DOAR database. Examples of the former may be possible repository saturation in certain countries and rationalization of multiple repositories into single systems (both discussed below). An example of the latter is the systematic deletion of dead repository records, a project that took place in accelerated form during 2012.

Geographical Distribution

Figure 2 illustrates the growth of repositories by continent between December 2005 and December 2012. *Open*DOAR uses ISO-3166 country attributions published by the International Organization for Standardization (ISO, n.d.) and associates them with nine continent designations: Africa, Asia, Europe, North America, Oceania, South America, and also Australasia, the Caribbean, and Central America (these last three areas are sometimes controversially subsumed within larger continents but are kept separate in *Open*DOAR).

The data show that Europe has consistently had a larger number of repositories than other continents, with the gap growing between it and North America (which shows shallower growth) during the period. Asia has shown a steeper growth pattern since late 2010, albeit at a lower level than Europe or North America. Africa has experienced low-level growth throughout the period. Unusually, Australasia is notable for having experienced a decline in the number of repositories since the beginning of 2011 (from 81 in January 2011 to 60 in December 2012). This is likely to be attributable to the closure of the Australasian Digital Theses (ADT) program in 2010, resulting in Australian universities merging their theses collections with their main IRs during the following year (Darby & Dyer, 2011).

More detail is shown in Figure 3, which illustrates the growth of repositories in the top 10 countries (in terms of numbers of repositories as at December 2012). The data show the United States to have had the largest number of repositories for the entire period covered by *Open*DOAR, followed by the United Kingdom and Germany. Germany's growth trend has been flatter than the United Kingdom's, although the U.K.'s growth has plateaued since the end of 2011. This plateauing raises the important issue of saturation, particularly for institutional repositories. It may be argued, for example, that with 209 repositories, most of which are institutional (see below), the United Kingdom is reaching a point where it is unlikely there will be significant further growth in the number of repositories because this figure reflects closely the number of major higher education and research institutions in the country.

Compared with countries such as Spain, France, and Italy, which have seen steady growth throughout the period covered by *Open*DOAR, other countries in the top 10 have seen an acceleration of growth later in the period, notably Taiwan (since mid-2010) and Brazil and Poland (since the beginning of 2011). Japan has clearly experienced growth from an earlier period (from 2006) onward; however, it follows a somewhat anomalous growth pattern, including two major step changes in growth. This is an artifact based on the data collection method for this and a small number of other non-English-language-speaking countries, which are derived from available English-language listings of repositories. In the case of Japan, data in *Open*DOAR are derived from the Japanese National Institute of Informatics listing of repositories (NII, n.d.) and reflect at least to some extent their update patterns.

A number of countries show relatively low levels of growth in repository numbers during the period studied. For example, China, in December 2006, was recorded as having two repositories; by December 2012, this number had risen

DOI: 10.1002/asi

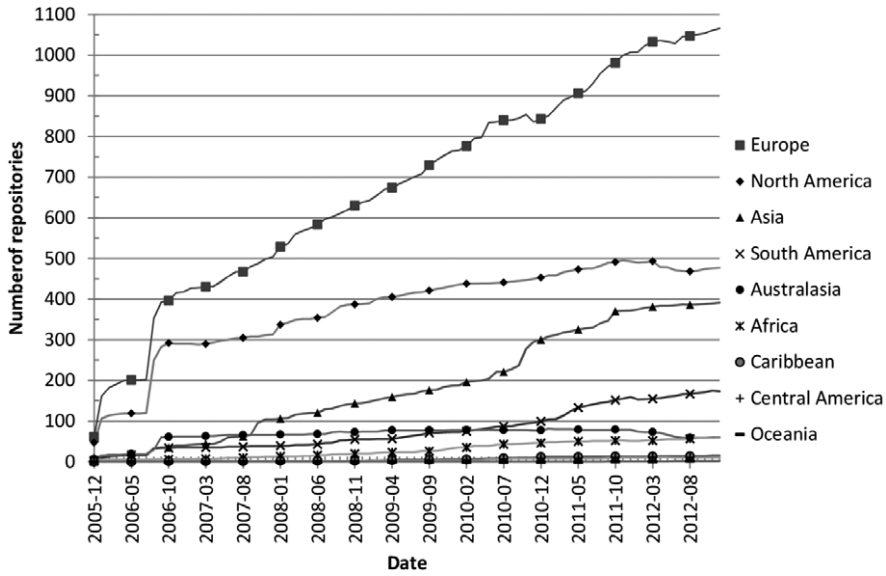

FIG. 2. Growth of repositories by continent, December 2005 to December 2012.

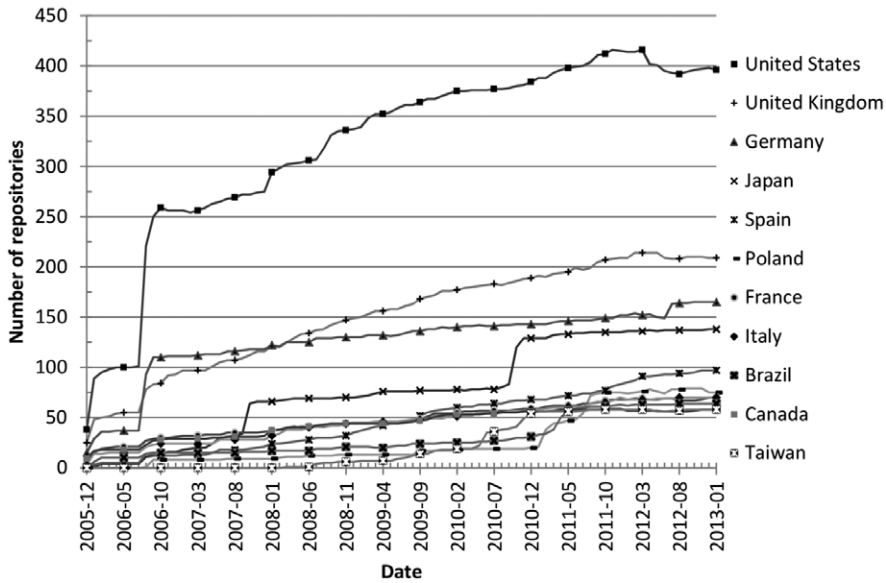

FIG. 3. Growth of repositories by country (top 10).

to only 33. During the same period, repository numbers in the Russian Federation rose from 1 to just 12.

Overall, between December 2006 and December 2012, the number of different countries in the *Open*DOAR database rose from 48 to 102. Tables 1 and 2 provide a comparison between the top 10 countries by repository numbers in December 2006 and December 2012. Table 1 shows the top 10 countries in 2006 and is comprised of countries in North America, Western Europe, plus Australia and Japan. Table 2 shows the 2012 top 11 countries (Canada and Taiwan are equal tenth) and illustrates the emergence of countries from Eastern Europe (Poland), South America (Brazil), and Asia

TABLE 1. Top 10 countries by repository numbers, December 2006.

Country	Number of repositories
United States	256
Germany	111
United Kingdom	93
Australia	52
Netherlands	44
France	31
Canada	29
Sweden	29
Italy	24
Japan	16

DOI: 10.1002/asi

TABLE 2. Top 11 countries by repository numbers, December 2012.

Country	Number of repositories
United States	398
United Kingdom	209
Germany	165
Japan	138
Spain	97
Poland	75
France	70
Italy	70
Brazil	64
Canada	58
Taiwan	58

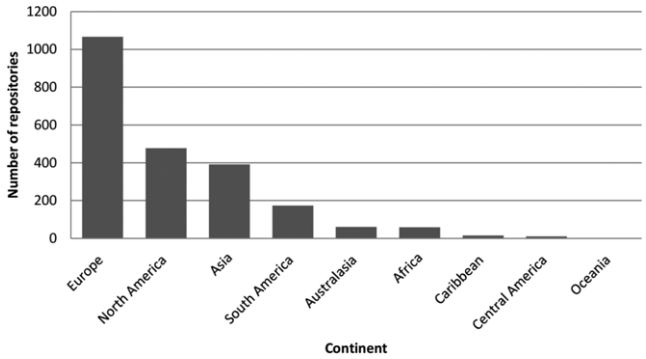

FIG. 4. Repositories by continent, December 2012.

(Taiwan). Also Spain, not previously in the top 10, has developed a large number of repositories compared with many other countries in Western Europe. France and Italy have remained in the top 10 with consistent development over the period. The Netherlands was initially in the top 10, but with an early repository development program and only a relatively small number of higher education institutions, seems to have reached saturation point early and therefore has been overtaken by countries with larger numbers of institutions.

An interesting feature between the two tables is that the number of repositories in the top country in 2006 (United States, $n = 256$) is 16 times that of the number of repositories in the tenth country (Japan, $n = 16$), whereas in 2012 this difference in the number of repositories between the top country and the tenth country was less than sevenfold (6.98). This is indicative of the growth in repository numbers outside of the relatively small group of early adopters.

The longitudinal data are complemented by the data collected on repositories in December 2012 derived from a snapshot of the entire *Open*DOAR database at that time. These show in detail the distribution of repositories by continent and country. Figure 4 presents the data by continent, showing Europe to have the largest number of repositories (1066; 47%), followed by North America (477; 21%), Asia (389; 17%), and South America (191; 8%).

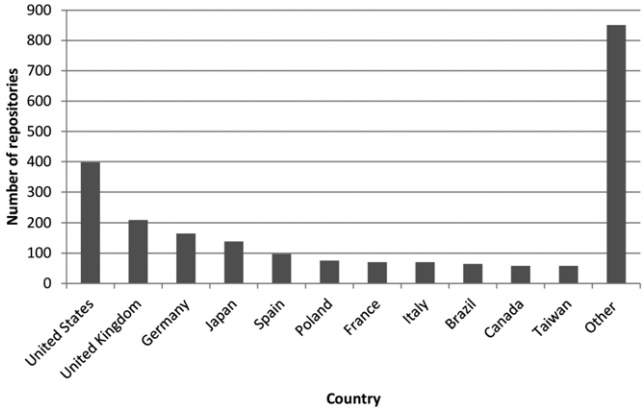

FIG. 5. Repositories by country, December 2012.

Figure 5 shows the extent to which the United States has a noticeably larger number of repositories than other countries (398; 18% of the total). The United States is followed by the United Kingdom (209; 9%), Germany (165; 7%), and Japan (138; 6%). These and the other top 11 countries (with Canada and Taiwan equal tenth with 58 repositories each) are responsible for a total of 1,402 repositories, 62% of all repositories worldwide.

The data therefore show a pattern of repository development initially concentrated in North America, Western Europe, and Australasia in the middle of the first decade of the 21st century, with Japan also showing significant development soon after. Growth continued in those locations but also occurred in other areas, notably Eastern Europe (particularly Poland), South America (particularly Brazil), and East Asia (particularly Taiwan). The worldwide distribution of repositories in December 2012 is illustrated in Figure 6 based on data where the *Open*DOAR record includes coordinates (present in 2,212 of the 2,253 records).

Repository Characteristics

A number of key characteristics of the repository infrastructure in December 2012 can be identified from the *Open*DOAR data. These include repository "type" (institutional, disciplinary, etc.), subject coverage (multidisciplinary or subject specific), content type (journal articles, theses etc.), language, and size. These data can be compared with those available from February 2008, which include a similar level of detail on repository characteristics, to identify any major changes since that time.

With regard to repository type, the vast majority of repositories worldwide are classified as "institutional" by *Open*DOAR (Figure 7). Institutional repositories account for 1,864 of the global 2,253 repositories, 83% of the total. Disciplinary repositories are the other major category, but these still only account for 11% (238) of the total. There were also 96 aggregating repositories (4% of the total) and 55 governmental repositories (2% of the total). These proportions have changed little since February 2008. Then, of

DOI: 10.1002/asi

FIG. 6. Map of repositories worldwide, December 2012. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

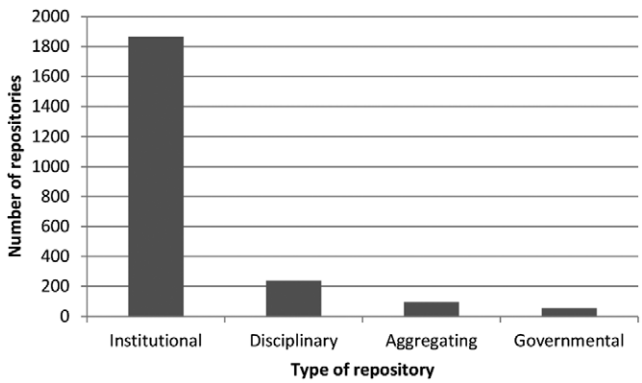

FIG. 7. Repository types, December 2012.

the 1,074 repositories, 860 (80%) were institutional, and 143 (13%) disciplinary. In addition, there were 5% aggregating and 2% governmental repositories.

A total of 1,857 different organizations are listed in the database as supporting repositories in December 2012. Data presented by Björk et al. (2013) suggest that these organizations include a large proportion of the "most productive" research institutions worldwide by publications. Using *Open*DOAR and ROAR data on institutional repositories, they made comparisons with the global league table of institutions produced by SCImago. They concluded, "Of the 148 most productive institutions, 82% had at least one institutional repository for the purpose of collecting, preserving and disseminating the intellectual output of the institution. The count of scientific documents for those institutions amounts to 3,620,234 (85%), which also gives some hints on the mass of authors given the possibility to self-archive in their affiliated institution's repositories." (Björk et al., 2013, p. 21). This analysis illustrates the potential of the repository infrastructure now established worldwide.

Most of the repositories are multidisciplinary in coverage, something which is to be expected if they are institutionally focused. Table 3 shows multidisciplinary repositories represent 1,390 (62%) of the repositories worldwide. Where repositories have a specific identifiable subject coverage (rather than being multidisciplinary), the subjects are recorded in *Open*DOAR. There are a wide range of subjects listed without any major subject area dominating: all are below 10% of the total number of repositories. Multidisciplinary repositories have increased their share of the overall repository infrastructure since February 2008. Then multidisciplinary repositories constituted 54% of the whole (581 of the 1,074), although similarly, no other subject category exceeded 10%.

As Figure 8 shows, the most common content type for repositories listed in *Open*DOAR is journal articles: 1,509 (67%) of repositories include this content type. Other content types are unpublished reports and working papers, including pre-refereed versions of articles (in 802; 36% of repositories), conference and workshop papers (783; 35%), and books, chapters, and sections of works (619; 27%). There are a number of other types, such as specific formats including learning objects and datasets, in a smaller number of repositories. Most repositories then include several

DOI: 10.1002/asi

TABLE 3. Repositories by subject, December 2012.

Subject	Frequency	Percentage
Multidisciplinary	1390	61.7
Health and medicine	210	9.3
History and archaeology	179	7.9
Business and economics	149	6.6
Law and politics	146	6.5
Technology general	145	6.4
Science general	144	6.4
Computers and IT	124	5.5
Education	123	5.5
Geography and regional studies	123	5.5
Social sciences general	119	5.3
Biology and biochemistry	112	5
Ecology and environment	107	4.7
Library and information science	97	4.3
Arts and humanities general	95	4.2
Agriculture, food and veterinary	86	3.8
Philosophy and religion	85	3.8
Language and literature	84	3.7
Mathematics and statistics	77	3.4
Fine and performing arts	75	3.3
Management and planning	73	3.2
Physics and astronomy	70	3.1
Chemistry and chemical technology	69	3.1
Psychology	54	2.4
Earth and planetary sciences	45	2
Mechanical engineering and materials	44	2
Architecture	36	1.6
Electrical and electronic engineering	35	1.6
Civil engineering	27	1.2

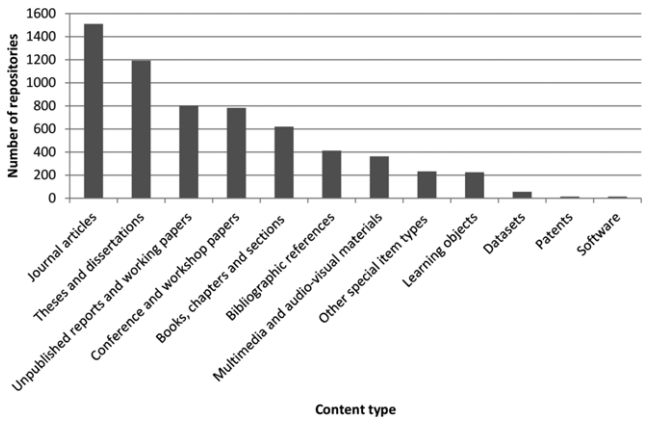

FIG. 8. Repository content types, December 2012.

content types that clearly support the scholarly communication process in various ways. These proportions have not markedly changed since 2008.

Language is also an important characteristic of repositories identified in *Open*DOAR. In December 2012, *Open*DOAR listed repositories containing resources in 57 languages. English was the most common language. 1,608 repositories (71%) included English-language material (Table 4). The predominance of English partly reflects the native-language of host countries of repositories but also no doubt reflects the position of the English language as the *lingua franca* of scholarship particularly in the science, technology and medicine domains (Tardy, 2004). There are, however, other common languages represented in repositories, also reflecting the countries of origin: Spanish (263, 12%), German (196, 9%), and Japanese (140, 6%). There are materials in a large number of languages included in a small number of repositories (10 or fewer). Repositories containing non-English language materials have increased as a proportion of the whole since February 2008. Of the 1,072 repositories on *Open*DOAR in February 2008, 85% (917) of them included English compared with 71% of the total in December 2012 evidencing repository development outside of English-speaking countries in the time since early 2008.

TABLE 4. Top 10 languages in repositories worldwide, December 2012.

Language	Frequency	Percentage
English	1608	71.4
Spanish	263	11.7
German	196	8.7
Japanese	140	6.2
French	129	5.7
Portuguese	114	5.1
Chinese	97	4.3
Polish	74	3.3
Italian	72	3.2
Norwegian	49	2.2

A particularly controversial issue associated with repositories is their size. The largest repositories contain millions of separate items. For example, one of the largest repositories in *Open*DOAR, Europe PubMed Central, contains 29,100,000. However, the vast majority of repositories are relatively small. Of the 2,129 repositories in *Open*DOAR with "size" data recorded, only 74 of them have more than 100,000 items, and 570, more than 10,000 items. Although the mean average of items in repositories is 66,526, the median number is only 3,093. Figure 9 shows a histogram for the size of repositories in *Open*DOAR on a logarithmic scale and Figure 10 shows the same data according to rank order of size. These illustrate the distribution of repository size across the entire population. The data show, in simple terms, that there is a small number of large repositories (i.e., with more than a million items) and a large number of small repositories (with between 100 and 100,000 items) that make up the global repository landscape. There is also a low number of very small repositories, with less that 6.6% of the population having 100 or fewer items. This is not surprising given that initial repository deployment may often tend to be associated with deposit of an early tranche of materials and organizations may not invest in initiating repository deployment without such a corpus being available.

DOI: 10.1002/asi

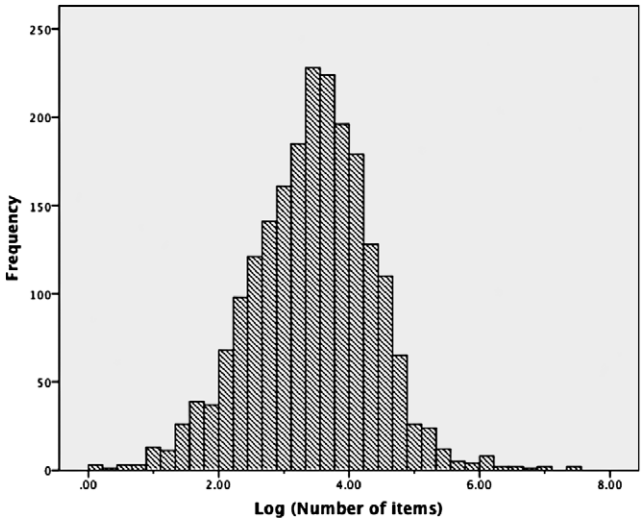

FIG. 9. Histogram of repository size, December 2012, on a logarithmic scale.

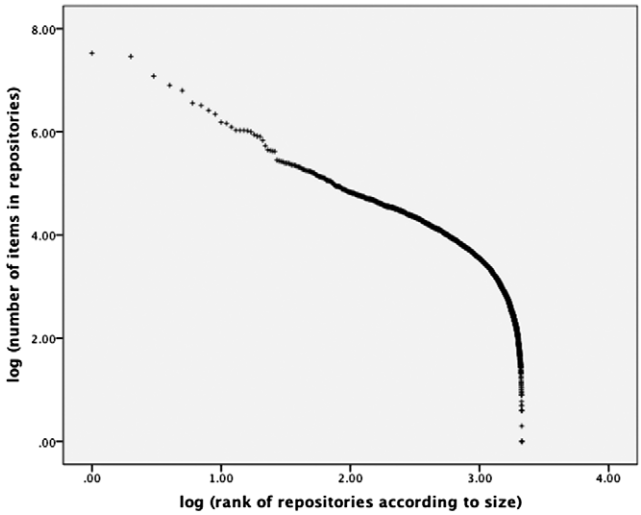

FIG. 11. Repository size according to rank, December 2012, on logarithmic scales.

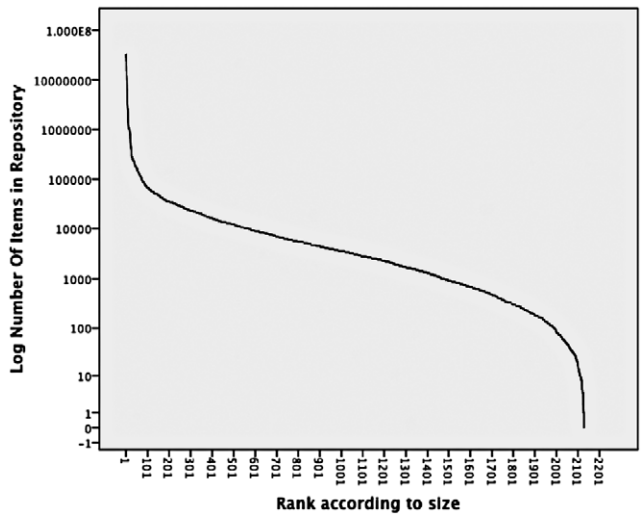

FIG. 10. Repository size, December 2012, on a logarithmic scale, according to rank order.

Figure 11 shows the number of items contained in repositories on a logarithmic scale against the log rank of the repositories and demonstrates a power law distribution over the bulk of the *x*-axis, with the relationship breaking down for the lower ranked repositories.

These data (and those in comparable services, such as ROAR) are, however, problematic. The data are gathered by the automated harvesting described previously, but crucially harvesters cannot distinguish between records in repositories with full content associated with them and those that are metadata only. There is no reliable way automatically to calculate only the number of full-content items housed in repositories, something that can only normally be achieved through case-by-case observation. For example, although Europe PubMed Central has more than 29 million items listed in *Open*DOAR, its own website states that it has more than 2.6 million full text items and more than 28 million abstracts. Therefore, approximately 10% of items in the repository are full text. The *Size* count in *Open*DOAR (or other directories) cannot distinguish between these full-text documents and the abstracts. In a case study of an institutional repository, Brown (2012) discussed the University of Southampton in the United Kingdom (a leader in institutional development) and reported that of the 17,317 deposits made to its repository in 2010, 12% were full text. This was an increase on previous years. The balance between full text and metadata records will, of course, vary across different repositories with different policies. Southampton, for example, uses its repository to collect metadata records of all of its publications as a matter of policy, encouraging the additional deposit of full text where possible. Other institutional repositories may only accept full text.

The repository size date in *Open*DOAR should therefore be considered maximum figures for full-content items in repositories. Although the number of full-content items in any repository may be lower than the *Open*DOAR size figure, it will not be higher, and therefore the *Open*DOAR figure represents an upper limit. This helps to put the median repository size figure of 3,093 into perspective because it is likely to be considerably higher than the actual number of full-content items. This emphasizes the point that to date repositories tend to be small. It also serves to highlight some significant consequences for assessments of repository development worldwide. It is common for researchers and OA advocates to make use of *Open*DOAR data (and those from similar services) uncritically in their analysis of repository size. This may mean that the overall size of the repository infrastructure is at times exaggerated. As observed by Björk et al. (2013), a variety of methods are needed to assess

DOI: 10.1002/asi

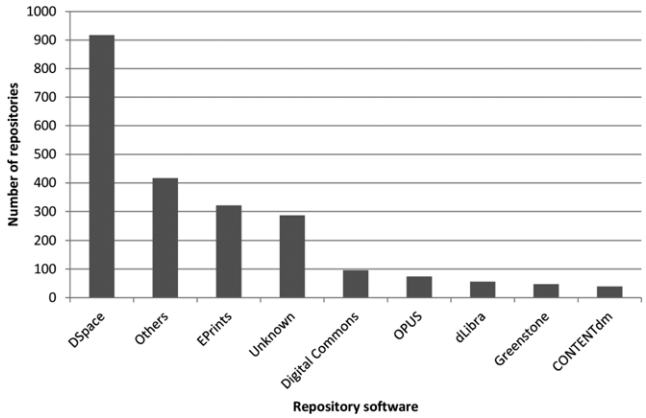

FIG. 12. Repository software types, December 2012.

repository size reliably, with *Open*DOAR data (and those of similar directories) being only one indicator.

Repository Setup

Repository providers typically make use of available open-source OAI-compliant software to run services, with two repository packages dominating: DSpace (917 repositories; 41%) and ePrints (321; 14%) (Figure 12). Between them, DSpace and ePrints account for over half of the repositories worldwide. These two software packages were also predominant in February 2008. However their relative positions have changed since that time. Of the 1,072 repositories then, 273 (25%) made use of DSpace, and 209 (19%) ePrints. DSpace has then noticeably gained market share.

Other repository software and platforms used include Digital Commons, OPUS and dLibra. OPUS use is limited to Germany and dLibra to Poland because these systems were designed in German and Polish respectively. In other non-English-speaking countries, there does not appear to be any language-specific software in widespread use, and so use is made of open-source software such as DSpace or locally developed solutions that cannot easily be identified. A significant minority of repositories (417 or 19%) use other software, and in the case of 286 (13%) the software could not be ascertained. A large proportion of repositories in these categories are likely to use locally developed software.

One key feature designed into many repository services is systemic interoperability. The *Open*DOAR project records where a repository has machine-to-machine protocols enabled (although these data are not yet publicly available in the directory). Unsurprisingly, the most common protocol being used by repositories (1,599; 71%) is OAI PMH, although this might be expected to be higher because most of the available open-source software repository packages come with OAI PMH "out of the box" and OAI PMH is the recognized protocol associated with the repository infrastructure. Other protocols being used are RSS (RSS 2: 943 repositories, 42%; RSS 1: 706, 31%; RSS—unspecified: 413, 18%), Atom (619; 27%) and SWORD (448; 20%).

As well as software used, *Open*DOAR also records repository license conditions. Strikingly, however, the 2,253 repositories listed in *Open*DOAR, as many as 1,766 have no evidence of licenses including rights information. Some repositories use licenses to define how content may be reused or harvested, for example, but they are a small minority. The absence of such licenses in the case of most repositories is a major weakness in the repository infrastructure. It means that content from a strictly legal point of view can only be assumed to be Gratis rather than Libre OA and therefore defined reuse rights are limited.

Discussion

The data presented here show growth in repositories worldwide between 2005 and 2012. Growth was concentrated initially in Western Europe, North America, and Australasia, closely followed by Japan. More recently, however, growth has widened to include other areas, particularly parts of Asia, Eastern Europe, and South America. Africa and Central Asia continue to experience comparatively low levels of repository development. The largest numbers of repositories are still in Europe, North America, and Australasia plus Japan. Repositories are predominantly institutional in their focus and multidisciplinary in their coverage. They contain a large number of content types, most of which are various forms of research outputs, and most commonly include English-language material. Although the size of repositories is difficult to assess accurately, the available data indicate that there is a small number of large repositories and a large number of small repositories making up the repository infrastructure. The most common technical setup for a repository involves the use of open-source OAI-compliant software but, significantly, often does not include any formal licensing of content.

Xia (2012) has demonstrated the use of IDT in explaining some of these global trends and characteristics. In particular, he makes some important observations about geographical variations in adoption of OA that apply to repositories as well as OA journals. He identifies a number of enablers and barriers to adoption that help to explain many of the key trends. In some cases, adoption is explained partly by different levels of existing IT infrastructure. Low levels of infrastructure provision in areas such as Africa and Central Asia, for example, may reasonably be assumed to be significant barriers to repository development. Language is also an important issue. In some cases the predominance of English in OA repositories may be a barrier in non-English-speaking countries, especially in disciplines where there may be a preference for communicating in other languages or a tendency, as observed by Norris, Oppenheim, and Rowland (2008), to "cite work from their own country preferentially." In other cases, language can be an enabler: with English increasingly being

DOI: 10.1002/asi

the "international language of science" (Garfield, 1977; see also Tardy, 2004), it may in many cases encourage communication across borders.

Cultural issues are also important in helping to explain significant variations in adoption between countries. Xia (2012) contrasts Japan and China, for example. In Japan, there was a preexisting scholarly communication culture that promoted sharing and which therefore created a climate in which rapid adoption of OA was possible. In contrast, in China, Xia argues that researchers, reacting to a widely perceived "gray area between reference and plagiarism," have shown caution about intellectual property loss and therefore a reluctance to adopt OA. In addition, policy developments have played a significant role. In the United Kingdom, Poland, and Brazil, a combination of OA policy developments plus funded initiatives from government-sponsored agencies encouraging OA adoption in general and repository development in particular have undoubtedly been important in facilitating growth in repository numbers. However, different policy approaches have led to different outcomes. For example, different development patterns in the United Kingdom and France have partly been the result of different policy approaches. Whereas the United Kingdom has encouraged and funded the development of distributed institution-based repositories largely through Jisc infrastructure programs, France has encouraged development of a smaller number of more centralized repositories, particularly the CNRS-funded (Centre National de la Recherche Scientifique) HAL (Hyper Articles en Lignes) series of repositories.

In this context, it is important to note that in China and Russia, where there are relatively low numbers of repositories, most of the repositories in existence are institutional. This means that the low number of repositories cannot be explained by a more centralized approach to repository development. In both cases then a combination of factors, including existing IT infrastructure, language, culture, policy, and funding, should be brought to bear to explain the current situation.

IDT as proposed by Rogers (2003) shows that in the case of successful innovations early adoption of the innovation is followed by accelerated adoption more widely, creating a classic S-shaped adoption curve associated with take-off. It is possible that OA repositories may follow this pattern of adoption globally particularly if repositories become an accepted part of mainstream scholarly practice. However, it is clear that this will partly depend on the precise role that repositories play in future. Advocates of Green OA see repositories as central to a future scholarly communication system, whereas others see repositories as playing a more peripheral role, for example, focusing on managing information such as grey literature and dissertations. The former is more likely to result in global take off, whereas the latter is more likely to result in incremental adoption.

Although it is reasonable to hypothesize that the use of repositories in a role central to the scholarly communication process in early-adopter countries may accelerate diffusion in other countries, it should be noted that repository development is likely to continue to be at different stages of the innovation cycle in different countries for the foreseeable future. Although there is considerable potential for continuing expansion of the repository infrastructure in many countries, it is probable that the numbers of repositories in certain (early adopter) countries may be nearing saturation point. For example, in the Netherlands and the United Kingdom, all major research institutions now have repositories, and it is therefore likely that growth in the number of repositories will plateau—data presented here suggest this may have begun to happen. This, of course, applies to the growth in the number of repositories rather than growth in their contents. As has been shown, for the vast majority of repositories, considerable potential remains to expand content—and IDT can again be usefully deployed to analyze such developments.

Jones et al. (2006) have also used IDT to explain repository development but at a different level from Xia. They use diffusion theory for framing a discussion on institutional repository development, and in particular the design of an advocacy program within an institution aiming to promote repository use amongst academic staff and therefore facilitate growth of repository content. Their analysis provides useful information for practitioners at institutional level while at the same time demonstrating the explanatory efficacy of IDT. They draw specifically on the explanation provided by Rogers (2003) of variables determining the rate of adoption of innovations. Rogers identifies five key attributes of innovations that significantly influence rates of adoption, described by Jones et al. as:

- "the *relative advantage* of adopting the new technology;
- the *compatibility* of existing work practices and ethics;
- the *complexity* involved in actually using the innovation;
- the *trialability*, or the degree to which it can be experimented with;
- the *observability* and visibility of the results."

(Jones et al., 2006, p. 114)

They go on to explain how IRs containing research outputs correspond to these attributes from the perspective of researchers in an institution. Specifically, they argue that an IR potentially delivers relative advantage to researchers by improving the economics of scholarly communication and increasing the visibility of their work. Use of the repository is compatible with existing academic values and requirements. It is relatively un-complex to use and can be trialed by academic staff. Finally, they argue, the repository's value can be observed through increased citation levels associated with OA and other visible impacts. Although these attributes ostensibly create conditions for rapid innovation uptake, major barriers remain, particularly around slow cultural acceptance of repositories and low levels of incentives encouraging individual adoption. Jones et al. suggest a number of practical approaches to addressing these barriers largely associated with awareness-raising and advocacy.

Jones et al.'s (2006) analysis relates primarily to the response of individuals to the innovation attributes proposed

DOI: 10.1002/asi

by Rogers (2003) and provides some useful commentary at that level. However, it is worth observing that the attributes may also reasonably be extended to apply to organizations and communities making innovation adoption decisions around setting up repositories in the first place as well as to individuals considering contributing content. In the case of repositories, relevant organizations and communities are academic institutions and disciplinary communities. With reference to Rogers' attributes of innovations, there is certainly an argument that OA repositories containing research outputs can deliver relative advantage to organizations by making the outputs of their staff more visible and therefore increasing academic and societal impact. Such an advantage can also be argued to be compatible with an academic institution's mission or that of a subject community. Furthermore, repositories are relatively simple to set up from an organizational point of view (although the embedding of business processes associated with self-archiving may be more complex). The use of repositories is also clearly trialable for organizations in a way that minimizes risk before widespread use is pursued. Finally, the results of repository use can be easily observed and monitored by organizations. These innovation attributes then seem to indicate significant potential for organizations in adopting repositories and provide a helpful framework for exploring many of the trends evident in the data reported in this study.

This application of IDT's innovation attributes to organizations complements other aspects of diffusion theory as proposed by Rogers (2003) that highlight the importance of organizations. Because organizational adoption of repositories is in most cases the primary innovation decision, with individual user adoption being contingent on that decision, organizational innovativeness is crucial. This is influenced by a number of variables identified by Rogers. The first of these is the characteristics of leaders in the organization and their attitude towards change. This has proved to be important in repository development with institutional and subject community leaders often being responsible for championing repository implementation at different stages. Certainly, the buy-in of senior managers in an institution and thought leaders in a community has been recognized to be an important factor in repository development (Jones et al., 2006). Much of the future of repository development, it may reasonably be assumed, is likely to be significantly determined by the role taken by senior leaders, especially as repository use widens and starts to come within the purview of a wider range of leaders in institutions not previously involved in OA development or advocacy.

Also important in determining organizational innovativeness are attributes of the organization itself as defined by Rogers (2003), including:

- Centralization: "the degree to which power and control in the system are concentrated in the hands of relatively few individuals"
- Complexity: "the degree to which an organization's members possess a relatively high level of knowledge and expertise"
- Formalization: "the degree to which an organization emphasizes following rules and procedures"
- Interconnectedness: "the degree to which the units in a social system are linked by interpersonal networks"
- Organizational slack: "the degree to which uncommitted resources are available"
- Size: with larger organizations often having greater capacity for innovation.

(Rogers, 2003, pp. 379–383)

Academic institutions and subject communities tend to have characteristics of low centralization, high complexity, and low formalization, with variable interconnectedness, variable organizational slack and variable size. These characteristics are particularly important when considering Rogers' distinction between different "subprocesses of the innovation": the initiation phase and the implementation phase. "Low centralization, high complexity, and low formalization facilitate initiation in the innovation process, but these structural characteristics make it difficult for an organization to implement an innovation." (Rogers, 2003, p. 380). This distinction between initiation on one hand and implementation on the other is highly relevant in considering organizational adoption of OA repositories to date. In many higher education institutions and subject communities, there may be few barriers to the initiation of innovation and as a result repositories can be set up quickly. However, achieving widespread implementation of the innovation in the organization is demonstrably more difficult, and the difficulty arises because of the same characteristics of the organization that made initiation relatively easy. The decentralized nature of decision making and processes, the high complexity of the organization and its governance, and the low formalization of structures and cultures, mean that the widespread implementation of the innovation, particularly in large organizations, can be delayed. There is then a complex relationship in terms of innovation adoption between the organization as a whole and individual members within it, with initial adoption of the innovation by the organization not necessarily resulting in short-term widespread acceptance by its members. In many early-adopter organizations, it may be observed, this is the stage their repositories have now reached, hence the concerns expressed in the practitioner literature on this topic.

Of course, users within any organization are not homogenous. There may be significant differences between users in terms of adoption of repositories, particularly associated with disciplinary differences (Björk et al., 2013; Kling & McKim, 2000). These have undoubtedly affected the take-up of disciplinary repositories, with some particular subject communities making extensive use of them (for example the high-energy physics community using arXiv). However, disciplinary differences also impact on the use of institutional repositories across a single organization (Spezi, Fry, Creaser, Probets, & White, 2013; Xia, 2008). Disciplinary differences therefore contribute to the complex set of

DOI: 10.1002/asi

factors (technological, cultural, and managerial) that affect the interrelationships between individuals, institutions, and scholarly communities, which have determined the shape of repository adoption to date and are likely to impact on the future development trajectories.

However, there is one additional factor to consider in this context, which has impacted on development to date and could potentially have a determining role in future adoption: the type of innovation decision. Rogers observes that decisions can either be optional (where individuals have choice in whether they adopt the innovation) or authority-based (where relatively small numbers of people in an organization or community determine adoption). In the case of many repositories, innovation initiation may have been carried out by a relatively small number of people but wider implementation often remains optional. Such optionality has often led to low levels of repository usage by faculty. However, IDT does recognize that on occasions "a higher level social organization such as a government, community, or a commercial company can exert its influence on the behavior of individual members of the system" and "mandate adoption" of the innovation (Rogers, 2003, p. 239). This often occurs where adoption can create system-wide benefits but the incentives for specific individuals to adopt voluntarily the innovation may be low—a factor that is important for OA in general and repository implementation in particular.

Mandates of various kinds have been introduced to encourage the adoption of OA, including use of repositories (Kennan, 2011; Xia et al., 2012). In most cases, mandates focus on author deposit of content, although imply, and may sometimes specify, organizational adoption of repositories on which user self-archiving is contingent. Some mandates have been introduced by institutions, encouraging OA adoption amongst their members (usually involving deposit in their own IR). However, the nature of the institutions as described above, has meant that these mandates are usually not easily enforceable, particularly in a context where incentives for implementation are often relatively low and sanctions for noncompliance minimal (Sale, 2006). There are, however, notable exceptions to this, where strongly worded policies combined with relevant incentives and sanctions can increase deposit rates significantly (Larivière et al., 2012). Nevertheless, as well as institutions, increasingly governments and research funders (government-sponsored and independent) have begun to introduce OA mandates. Because their mandates generally have a much wider coverage than institution-specific policies and also, because they often introduce meaningful incentives and sanctions, they are often more likely to create significant levels of change in author behavior, particularly when combined with institution-level compliance programs.

Mandates have been introduced by a number of funders since 2003. In that year, a number of research funders in Germany issued the Berlin Declaration announcing declaring their support for open access (Max Planck Society, 2003). A year later in the United States, the National Institutes of Health (NIH) in the United States introduced a mandate for all of its grant holders to make their work OA (O'Keeffe, Willinsky, & Maggio, 2011). This was followed in 2005 by the Wellcome Trust in the United Kingdom introducing a more strongly worded mandate (Walport & Kiley, 2006). A large number of other governments and funders have introduced such policies since. In the past 2 years, there has been and further activity to strengthen the wording of such mandates, "requiring" rather than "encouraging" deposit in a repository, for example. Also, the introduction of meaningful incentives and sanctions to encourage compliance has been apparent, including, for example, withholding parts of grants until compliance can be demonstrated. At the same time, levels of compliance are also beginning to be monitored by funders. In the United Kingdom, for example, the Wellcome Trust now actively monitors compliance with its mandate and has introduced measures involving sanctions, such as withholding funding in various ways in the event of noncompliance (Wellcome Trust, 2012). In January 2013, Kiley (2013) reported compliance levels for Wellcome-funded research outputs to be around 60% in March 2012 compared with less than 20% in March 2007. Also, the government-sponsored Research Councils have issued a joint policy that aims to ensure that all of its grant holders make their work open access, preferably via the Gold route but also allowing for Green OA (RCUK, 2013). More recently, the Higher Education Funding Council for England has proposed that the research assessment process that it manages (the Research Excellence Framework) will in future involve the expectation that all work submitted for assessment should be made available in an institutional repository (HEFCE, 2013). If the latter policy were to be implemented, it would significantly increase the likelihood of widespread adoption of repositories by UK researchers. Similar developments are underway, for example, in the United States, with the Federal Fair Access to Science and Technology Research (FASTR) bill announced in February 2013 and a new OA policy memorandum issued by the Office of Science and Technology Policy (OSTP) the same month (Kaiser, 2013).

Evidence suggests that mandates may be crucial in determining the future development trajectory of repositories, particularly in terms of widespread implementation and use. If they are to be effective, mandates clearly need to be worded in a robust way and complemented by meaningful incentives and sanctions as well as accompanied by compliance monitoring. The importance of the policy developments now taking place particularly in the repository early-adopter countries, are yet to be seen both in terms of OA in general and repositories in particular. However, there is a strong possibility that they will have a significant impact on the use of the repository infrastructure in those countries and that impact will in turn accelerate diffusion in other countries, both in terms of the growth of repository numbers and repository content. It is likely, however, that different countries will continue to proceed at different speeds depending on the complex set of technical, cultural, managerial and economic factors. Such developments need to continue to be tracked and monitored.

DOI: 10.1002/asi

At the same time, cultures of communication are changing. Greater experimentation in scholarly communication is happening in a large number of ways, much of it increasingly open access. Although this varies across disciplines, and disciplinary differences will certainly continue to be important, it is likely to impact on the use of repositories by various subject communities. This is particularly the case as other open developments, including data sharing and exchange of learning objects (which are themselves often associated with repositories), also become more embedded in academic practice.

Conclusion

The data from the *Open*DOAR service presented here provide an important perspective on the development of the global repository infrastructure from 2005 to 2012. They show that early adoption, concentrated in Western Europe, North America, and Australasia, has been followed by development in other areas, including East Asia, South America, and Eastern Europe. However, areas such as Africa and Central Asia, and countries such as China and Russia, have experienced relatively low levels of growth throughout the period. Globally, repositories are predominantly institutional, multidisciplinary, and English-language-based. They are typically set up using open-source OAI-compliant repository software but remain immature in terms of explicit licensing arrangements. Crucially, they are for the most part relatively small in terms of content. Further work is needed to continue to monitor global development trends of repositories and to assess the importance of the various factors that influence those developments. In particular, more work is required to assess reliably content growth in repositories. Work is also required to strengthen repository services themselves, particularly around the creation of explicit license agreements to enable the reuse of resources.

The global trends described here can be explained by a number of factors informed particularly by IDT. Among these, existing IT infrastructure and language play an important role. However, a complex set of cultural factors along with policy and funding initiatives are also crucial. They suggest that different countries may continue to proceed at different speeds in the innovation process with the ultimate trajectory of the growth curve dependent on the final role the repositories have in scholarly practice.

The setting up of repositories by particular institutions or subject communities is partly determined by the attributes of the innovation itself. It may be seen that repositories create ostensible relative advantage for the organization adopting them and its members: They are compatible with existing working practices and values, they are relatively uncomplicated to use, they are trialable, and their results are visible—factors that create potential for successful innovation. As well as the innovation attributes, the attributes of the organization are also crucial. However, it may be observed that the attributes of academic organizations, which may make initiation of repository developments relatively easy (including low centralization, high complexity, and low formalization), may actually contribute to making widespread implementation of repositories problematic. Evidence suggests that many repositories, particularly institutional repositories, are not being widely used by researchers with overall deposit levels remaining relatively low. Strong leadership in an organization combined with awareness-raising and advocacy programs promoting the use of repositories (themselves informed by the key innovation and organizational attributes relating to repositories) will undoubtedly go some way to addressing these issues. Further work is needed to analyze the effectiveness of such initiatives.

However, an equally important consideration is the level of voluntary adoption of repositories. Evidence suggests that mandates may accelerate adoption and therefore may be crucial in determining future adoption patterns. Although institutional mandates have some role to play, funder or governmental mandates are more likely to have a greater impact. Their impact does, however, largely depend on the extent to which they are worded robustly, complemented by meaningful incentives and sanctions, and accompanied by compliance monitoring. Recent developments in these areas in a number of repository early-adopter countries are likely to have a significant impact within their own countries, and this is likely to extend more widely. Their impact needs to be monitored and analyzed.

The complex relationships between individual researchers, their organizations and subject communities, their funders and governments, and the contributions all these actors make to repository development need to be further explored to explain ongoing developments. OA approaches in general, and OA repositories in particular, have the potential to transform scholarly practice. The period since 2005 has seen considerable change in this area, with a repository infrastructure being established that is capable of playing an important role in scholarly communication. However, it is the next decade that is likely to reveal the extent to which these changes make a widespread and enduring impact on the scholarly community.

Acknowledgments

*Open*DOAR has received funding from Jisc, the Open Society Institute, Research Libraries UK, and SPARC Europe.

References

Bhat, M. (2010). Interoperability of open access repositories in computer science and IT—An evaluation. Library Hi Tech, 28(1), 107–118. doi:10.1108/07378831011026724

Björk, B., Laakso, M., Welling, P., & Paetau, P. (2013). Anatomy of green open access. Journal of the American Society for Information Science and Technology, (Pre-print). Retrieved from http://www.openaccesspublishing.org/apc8/Personal%20VersionGreenOa.pdf

Brown, M. (2012). Open access at the University of Southampton—Pushing the boundaries and the art of the impossible. Jisc. Retrieved from http://www.jisc.ac.uk/whatwedo/topics/opentechnologies/openaccess/case-studies/southampton-university.aspx

Buehler, M. (2013). Demystifying the Institutional Repository for Success. Cambridge: Woodhead Publishing.

Connell, T. (2011). The use of institutional repositories: The Ohio State University experience. College & Research Libraries, 72(3), 253–275. Retrieved from http://crl.acrl.org/content/72/3/253.short

Crow, R. (2002). The case for institutional repositories: A SPARC position paper. ARL Bimonthly Report 223. Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1006&context=ir_research

Darby, N., & Dyer, K. (2011). ADT—Original Australian research in the spotlight. Libraries Australia. Retrieved from http://www.nla.gov.au/librariesaustralia/files/2011/07/issue-11.pdf

Davis, P., & Connolly, M. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. D-Lib Magazine, 13(3/4). doi:10.1045/march2007-davis

Dulle, F.W., & Minishi-Majanja, M.K. (2011). The suitability of the Unified Theory of Acceptance and Use of Technology (UTAUT) model in open access adoption studies. Information Development, 27(1), 32–45. doi:10.1177/0266666910385375

Foster, N., & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. D-Lib Magazine, 11(1). doi:10.1045/january2005-foster

Garfield, E. (1977). English—An international language for science. Information Scientist, 19–20.

Ginsparg, P. (1997). Winners and losers in the global research village. The Serials Librarian, 30(3–4), 83–95. doi:10.1300/J123v30n03_13

Gutteridge, C. (2002). GNU EPrints 2 overview. In Proceedings of the 11th Panhellenic Academic Libraries Conference. Retrieved from http://eprints.soton.ac.uk/256840/

Halpern, J., & Lagoze, C. (1999). The Computing Research Repository: Promoting the rapid dissemination and archiving of computer science research. Proceedings of the fourth ACM conference on Digital Libraries, 3–11. doi:10.1145/313238.313240

Hanlon, A., & Ramirez, M. (2011). Asking for permission: A survey of copyright workflows for institutional repositories. portal: Libraries and the Academy, 11(2), 683–702. Retrieved from http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v011/11.2.hanlon.html

Harnad, S. (n.d.). Open Access Archivangelism: Blog. Retrieved from http://openaccess.eprints.org/

HEFCE. (2013). Call for advice on open access. Retrieved from http://www.hefce.ac.uk/news/newsarchive/2013/name,78750,en.html

ISO. (n.d.). Country Codes—ISO 3166. Retrieved from http://www.iso.org/iso/country_codes

Jones, R., Andrew, T., & MacColl, J. (2006). The Institutional Repository. Oxford: Chandos Publishing.

Kaiser, J. (2013, February 22). White House unveils long-awaited public access policy. Science Magazine Online. Retrieved from http://news.sciencemag.org/scienceinsider/2013/02/white-house-unveils-long-awaited.html

Kennan, M.A. (2011). Learning to share: Mandates and open access. Library Management, 32(4/5), 302–318. doi:10.1108/01435121111132301

Kiley, R. (2013). Europe PMC: A gold and green OA repository. In 5th Couperin's Open Access meeting. Retrieved from http://couperin.sciencesconf.org/conference/couperin/pages/RobertKiley.pdf

Kim, J. (2010). Faculty self-archiving: Motivations and barriers. Journal of the American Society for Information Science and Technology, 61(9), 11909–1922. doi:10.1002/asi.21336

Kim, J. (2011). Motivations of faculty self-archiving in institutional repositories. The Journal of Academic Librarianship, 37(3), 2246–254. doi:10.1016/j.acalib.2011.02.017

Kling, R., & McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. Journal of the American Society for Information Science and Technology, 51(14), 1306–1320. doi:10.1002/1097-4571

Lagoze, C., & Sompel, H. Van de. (2003). The making of the Open Archives Initiative Protocol for Metadata Harvesting. Library Hi Tech, 21(2), 118–128. doi:10.1108/07378830310479776

Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2012). Testing the Finch Hypothesis on Green OA Mandate Ineffectiveness. arXiv. Retrieved from http://arxiv.org/abs/1210.8174

Lawal, I. (2002). Scholarly communication: The use and non-use of e-print archives for the dissemination of scientific information. Issues in Science and Technology Librarianship, 36. Retrieved from http://www.istl.org/02-fall/article3.html

Lynch, C. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. portal: Libraries and the Academy, 3(2), 327–336. doi:10.1353/pla.2003.0039

Mann, F., von Walter, B., Hess, T., & Wigand, R.T. (2009). Open access publishing in science. Communications of the ACM, 52(3), 135. doi:10.1145/1467247.1467279

Max Planck Society. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 2003. Retrieved from http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/

McKiernan, G. (2000). arXiv.org: The Los Alamos National Laboratory e-print server. International Journal on Grey Literature, 1(3), 127–133. doi:10.1108/14666180010345564

Millington, P. (2013). Counting on OpenDOAR. In Open Repositories, 8–11 July 2013. Retrieved from http://www.sherpa.ac.uk/documents/Counting-on-OpenDOAR-2013-07-10.pptx

Morrison, H. (2012). December 31, 2012 Dramatic Growth of Open Access. Retrieved from http://poeticeconomics.blogspot.co.uk/2012/12/december-31-2012-dramatic-growth-of.html

Nabe, J.A. (2009). Starting, strengthening and managing institutional repositories: A how-to-do-it manual. New York: Neal-Schuman.

Nazim, M., & Mukherjee, B. (2011). Status of institutional repositories in Asian countries: A quantitative study. Library Philosphy & Practice. Retrieved from http://digitalcommons.unl.edu/libphilprac/652

Nicholas, D., Rowlands, I., Watkinson, A., Brown, D., & Jamali, H.R. (2012). Digital repositories ten years on: What do scientific researchers think of them and how do they use them? Learned Publishing, 25(3), 195–206. doi:10.1087/20120306

NII. (n.d.). NII Institutional Repositories DataBase Contents Analysis. Retrieved from http://irdb.nii.ac.jp/analysis/browse_e.php

Norris, M., Oppenheim, C., & Rowland, F. (2008). Open access citation rates and Developing Countries. In ELPUB 2008 Conference on Electronic Publishing (pp. 335–342). Retrieved from http://elpub.scix.net/data/works/att/335_elpub2008.content.pdf

Nyambi, E., & Maynard, S. (2012). An investigation of institutional repositories in state universities in Zimbabwe. Information Development, 28(1), 55–67. doi:10.1177/0266666911425264

O'Keeffe, J., Willinsky, J., & Maggio, L. (2011). Public access and use of health research: an exploratory study of the National Institutes of Health (NIH) Public Access Policy using interviews and surveys of health personnel. Journal of Medical Internet Research, 13(4), e97. doi:10.2196/jmir.1827

Ochoa, X., & Duval, E. (2009). Quantitative analysis of learning object repositories. IEEE Transactions on Learning Technologies, 2(3), 226–238. doi:http://dx.doi.org/10.1109/TLT.2009.28

OpenDOAR. (n.d.-a). About OpenDOAR—Directory of Open Access Repositories. Retrieved from http://www.opendoar.org/about.html

OpenDOAR. (n.d.-b). The fate of suggestions 2008 to date. Retrieved from http://www.opendoar.org/suggestionschart.html

Palmer, C., Teffeau, L., & Newton, M. (2008). Strategies for institutional repository development: A case study of three evolving initiatives. Library Trends, 57(2), 142–167. doi:10.1353/lib.0.0033

Pinfield, S. (2001). How do physicists use an e-print archive? Implications for institutional e-print services. D-Lib Magazine, 7(12). doi:10.1045/december2001-pinfield

Pinfield, S. (2005). A mandate to self archive? The role of open access institutional repositories. Serials: The Journal for the Serials Community, 18(1), 30–34. doi:10.1629/1830

Pinfield, S. (2009). Journals and repositories: An evolving relationship? Learned Publishing, 22(3), 165–175. doi:10.1087/2009302

RCUK. (2013). RCUK Policy on Open Access and Supporting Guidance. Retrieved from http://www.rcuk.ac.uk/documents/documents/RCUKOpenAccessPolicy.pdf

Rogers, E.M. (2003). Diffusion of innovations (5th ed.). New York: Simon & Schuster International.

Sale, A. (2006). Comparison of content policies for institutional repositories in Australia. First Monday, 11(4). Retrieved from http://www.firstmonday.dk/ojs/index.php/fm/article/view/1324

Sawant, S. (2011). Institutional repositories in India: A preliminary study. Library Hi Tech News, 28(10), 6–10. doi:10.1108/07419051111202646

Shearer, K. (2006). The CARL institutional repositories project: A collaborative approach to addressing the challenges of IRs in Canada. Library Hi Tech, 24(2), 165–172. doi:10.1108/07378830610669547

Singeh, F., Abrizah, A., & Karim, N.H.A. (2013). Malaysian authors' acceptance to self-archiving institutional repositories: Towards a unified view. Electronic Library, 31(2), 188–207. doi:10.1108/02640471311312375

Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., . . . Walker, J.H. (2003). DSpace: An open source dynamic digital repository. D-Lib Magazine, 9(1). doi:10.1045/january2003-smith

Sompel, H. Van de, & Lagoze, C. (2000). The Santa Fe Convention of the Open Archives Initiative. D-Lib Magazine, 6(2). doi:10.1045/february2000-vandesompel-oai

Spezi, V., Fry, J., Creaser, C., Probets, S., & White, S. (2013). Researchers' green open access practice: A cross-disciplinary analysis. Journal of Documentation, 69(3), 3334–359. doi:10.1108/JD-01-2012-0008

Suber, P. (2008, August). Gratis and libre Open Access. SPARC Open Access Newsleter. Retrieved from http://www.sparc.arl.org/publications/articles/gratisandlibre.shtml

Suber, P. (2012). Open access. Boston, MA: MIT Press.

Tardy, C. (2004). The role of English in scientific communication: Lingua franca or Tyrannosaurus rex? Journal of English for Academic Purposes, 3(3), 247–269. doi:10.1016/j.jeap.2003.10.001

Tripathi, M., & Jeevan, V. (2011). An evaluation of digital libraries and institutional repositories in India. Journal of Academic Librarianship, 37(6), 543–545. doi:10.1016/j.acalib.2011.08.012

Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User acceptance of information technology: Toward a unified view. MIS quarterly, 27(3), 425–478. Retrieved from http://www.jstor.org/stable/10.2307/30036540

Walport, M., & Kiley, R. (2006). Open access, UK PubMed Central and the Wellcome Trust. Journal of the Royal Society of Medicine, 99(9), 438–9. doi:10.1258/jrsm.99.9.438

Walters, T. (2007). Reinventing the library: How repositories are causing librarians to rethink their professional roles. portal: Libraries and the Academy, 7(2), 213–225. doi:10.1353/pla.2007.0023

Warr, W.A. (2003). Evaluation of an experimental chemistry pre-print server. Journal of Chemical Information and Computer Sciences, 43, 362–373.

Wellcome Trust. (2012). Wellcome Trust strengthens its open access policy. Retrieved from http://www.wellcome.ac.uk/News/Media-office/Press-releases/2012/WTVM055745.htm#

Xia, J. (2007). Assessment of self-archiving in institutional repositories: Across disciplines. The Journal of Academic Librarianship, 33(6), 647–654. doi:10.1016/j.acalib.2007.09.020

Xia, J. (2008). A Comparison of Subject and Institutional Repositories in Self-archiving Practices. The Journal of Academic Librarianship, 34(6), 489–495. doi:10.1016/j.acalib.2008.09.016

Xia, J. (2012). Diffusionism and open access. Journal of Documentation, 68(1), 72–99. doi:10.1108/00220411211200338

Xia, J., Gilchrist, S., Smith, N., Kingery, J., Radecki, J., Wilhelm, M., . . . Mahn, A. (2012). A review of open access self-archiving mandate policies. portal: Libraries and the Academy, 12(1), 85–102. doi:10.1353/pla.2012.0000

Yiotis, K. (2008). Electronic theses and dissertation (ETD) repositories: What are they? Where do they come from? How do they work? OCLC Systems & Services, 24(2), 101–115. doi:10.1108/10650750810875458