Open Access Subject Repositories: An Overview

Bo-Christer Björk
Hanken School of Economics, P.O. Box 479, 00101 Helsinki, Finland. E-mail: bo-christer.bjork@hanken.fi

Subject repositories are open web collections of working papers or manuscript copies of published scholarly articles, specific to particular scientific disciplines. The first repositories emerged in the early 1990s, and in some fields of science they have become an important channel for the dissemination of research results. With quite strict inclusion criteria, 56 subject repositories were identified from a much larger number indexed in 2 repository indices. A closer study of these demonstrated a huge variety in sizes, organizational models, functions, and topics. When they first started to emerge, subject repositories catered to a strong market demand, but the later development of Internet search engines, the rapid growth of institutional repositories, and the tightening of journal publisher open access policies seems to be slowing their growth.

Introduction

The emergence of the Internet has radically enhanced the possibilities for scientists to disseminate their research ideas and publications directly to potential readers and to bypass the long time delays and selection processes required by traditional publishing. It costs nothing for a scholar to put a working paper up on the web and hope that others will read it, cite it, and provide links to it, as well as feedback. Nevertheless, outreach is usually not very good, unless the author happens to be a very well known scientists whose writings are followed by many.

Before the web, paper-based services for the dissemination of manuscript-stage publications had emerged in certain disciplines. In 1991 the first Internet-based subject repository, arXiv, emerged (Ginsparg, 2004). Such repositories offered the possibility for authors to embed their “papers” in a critical mass of other manuscripts on similar topics, which tends to attract more potential readers. This type of dissemination also provides some degree of quality assurance via the brands of the repositories as well as a safer and more stable storage place compared with personal or departmental web pages. Although subject repositories may also contain article metadata, like traditional citation indices, as well as research data, most of the interesting ones provide full texts of scholarly publications available free of charge and are searchable by web robots.

In a broader context, *subject repositories* provide one of a number of channels for providing scholarly *open access (OA) literature* (Suber, 2012). For the peer-reviewed journal literature, the journals themselves can be open access (often called “gold OA”). In 2010, over 8,000 such journals published approximately 340,000 articles (Laakso & Björk, 2012). This can be achieved by a number of alternative business models, with the model in which authors pay for the publication services rapidly becoming more common (Björk & Solomon, 2012). The other main alternative is author self-archiving of manuscript copies openly on the web (green OA), either on their own or on their department’s web pages, in the institutional repositories of their universities, or in subject repositories, which the topic of this study. In an earlier study, we found that 20% of the peer-reviewed articles published in 2008 were openly available, with green OA contributing 12% (Björk et al., 2010). Green OA subject repositories provided the channel for about one third of the articles (Björk, Laakso, Welling, & Paetau, 2013).

Literature Review

Quite a lot has been written about repositories in general (Armbruster & Romary, 2009; Kim, 2010), about author attitudes toward uploading green copies (Nicholas, Rowlands, Watkinson, Brown, & Jamali, 2012; Kleinman, 2011), and about the citation advantage of self-archiving in them (Swan, 2010; Wagner, 2010), but few studies have concentrated specifically on subject repositories. Most of these have presented case studies of successful repositories, often written by the scholars who developed them. In their systematic literature review of studies on this topic, Adamick and Reznik-Zellen (2010a) in fact state that “subject repositories are under-studied and under-represented in library science literature and in the scholarly communication and digital library fields” and further that

Received February 12, 2013; Revised May 14, 2013; Accepted May 15, 2013

© 2013 ASIS&T • Published online 14 November 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23021

"the lack of subject repository recognition within the literature . . . may be attributed to the isolated development of the largest subject repositories and a general lack of awareness about small-scale subject repositories." Adamick and Reznik-Zellen collected papers written after the year 2000 on the 10 biggest subject repositories and found only six articles discussing subject repositories more broadly, in contrast to 31 articles discussing individual repositories in rather practical terms.

Kling and McKim (2000) were the first to highlight how the differences in knowledge-sharing cultures between scholarly fields prior to the Internet could explain the success of early subject repositories in fields such as physics and economics. Darby, Jones, Gilbert, and Lambert (2008) looked at the interfaces between subject repositories and institutional repositories, whereas Xia (2008) compared the self-archiving behavior of physicists in both subject and institutional repositories. Several publications have highlighted how successful repositories have emerged, in particular the organizational structures that have allowed success (Parinov & Krichel, 2004; DeRobbio & Katzmayr, 2009; Ginsparg, 2004; Kelly & Letnes, 2006). Adamick and Reznik-Zellen (2010b) also followed the literature study mentioned earlier with an empirical study of the 10 biggest subject repositories, but little is known about the vast majority of smaller repositories. The main purpose of this study was thus to provide a broader understanding of subject repositories and their development, going beyond the few success stories and looking in particular at the range of organizational structures used as well as to provide a better understanding of the size distribution, topical range, services, country of origin, and information technology (IT) platforms used.

Methods

Hundreds of subject repositories are included among the more than 2,000 repositories listed in either the Directory of Open Access Repositories (DOAR) or the Registry of Open Access Repositories (ROAR). Also, in addition to the repositories indexed in these directories, there exists an unknown number of smaller repositories and failed attempts to build repositories.

The only practical way to select repositories for closer scrutiny was to start with the ones indexed in either DOAR or ROAR. Both of these classify repositories into a number of generic types, so it was simple to exclude institutional ones (the vast majority) from further investigation. Data on repositories in DOAR listed as either disciplinary (235) or aggregating (96) and in ROAR as research cross-institutional (226) were first collected on November 21, 2012. This led to an initial list of 503 candidates, including many duplicates for repositories included in both indices.

A cursory look at the candidate repositories revealed that a majority of them did not really fit the description usually given in the OA literature. With this in mind, criteria for inclusion in a shorter list were defined. Repositories that were within scope were such that:

- They have a clear subject limitation.
- There must be a channel for authors regardless of affiliation to upload a manuscript as long as it is within the topic area.
- At least part of the content consists of working papers and/or submitted or accepted articles.
- Access must be open, with no charges to most of the publications.

Types of repositories that were discarded included:

- Institutional repositories
- Multi-institution repositories with upload restricted to authors from member institutions only
- Repositories for members of particular associations or projects only
- "Orphan" repositories with no subject limitations (repositories meant for authors from institutions lacking an institutional repository)
- Repositories meant for masters and PhD theses only
- OA journal portals
- Conference proceedings portals
- Websites focusing on reusable teaching materials, books
- Historical document archives
- Repositories that charge authors
- Directories with only metadata
- Portals with only link lists
- Services no longer found, with broken links

In some cases it was difficult to draw a line, and a couple of border, but particularly interesting, repositories were included. After the analysis, 56 repositories remained, ranging from very large repositories with hundreds of thousands of documents to almost unpopulated ones. These were studied using the data available from the ROAR and DOAR indices, by going to their websites (in particular the "about" pages), and by searching the web for literature about them. The basic data on these repositories are given in Table 1.

The 56 chosen are not a random sample of all the repositories first extracted from the indexed repositories. From a methodological viewpoint, this study could instead be labeled a multicase study. On the other hand, the repositories in focus are the ones that fit our own definition of subject repositories oriented toward dissemination of journal articles and the working papers that precede them.

Results

Size Distribution

The data on the size of the repositories are not particularly accurate, particularly if we are interested only in articles and want to exclude other types of items. For the analysis, data from both DOAR and ROAR were used. The data on the number of documents are not well established; the larger number reported in either ROAR and DOAR was usually used, and in some cases the number was checked from the website. For some repositories it was also possible to use the browsing function directly. For all the repositories using the EPrints software, it was possible to extract the

TABLE 1. The 56 repositories studied.

	Range				Items	Founded	Country	Software	Type
Physics and mathematics									
arXiv	✓				805,000	1991	USA	In-house	Independent
viXra	✓				3,680	2009	UK	In-house	Independent
Economics and management									
Social Science Research Network	✓				814,725	1992	USA	In-house	Independent
Research Papers in Economics		✓			1,200,000	1993	USA	In-house	Independent
Munich Personal RePEc Archive		✓			22,643	2006	Germany	EPrints	Institution
Socionet		✓			3,520	2006	Russia	In-house	Independent
Econstor		✓			48,252	2009	Germany	DSpace	Institution
Industry Studies Working Papers				✓	130	2010	USA	EPrints	Association
Computing and information science									
CiteSeer^X	✓				2,000,000	1997	USA	In-house	Institution
E-LIS		✓			14,053	2003	Italy	DSpace	Association
Arab Repository for Libr. and Inf. Studies		✓			52	2010	Egypt	In-house	Institution
Archivesic		✓			1,501	2002	France	HAL	Institution
DLIST		✓			1,540	2002	USA	DSpace	Institution
Sprouts				✓	485	2000	USA	In-house	Association
Cryptology ePrint Archive				✓	5,702	1996	USA	In-house	Association
Architektur-Informatik				✓	113	2003	Austria	SciX	Independent
ERPAePRINTS				✓	82	2003	UK	EPrints	Institution
AgentLink				✓	1,410	2004	UK	EPrints	Association
Graph Drawing E-Print Archive				✓	886	2003	Germany	EPrints	Institution
Medicine									
PubMed Central	✓				2,600,000	2000	USA	In-house	Institution
OpenMED@NIC	✓				2,866	2005	India	In-house	Institution
Clinical Medicine NetPrints	✓				81	1999	UK	In-house	Association
Dryad	✓				8,849	2009	USA	DSpace	Institution
Philosophy									
PhilPapers		✓			507,277	2006	UK	In-house	Institution
PhilSci Archive			✓		3,005	2000	USA	EPrints	Institution
Sammelpunkt. Elektronisch archivierte Theorie	✓				1,526	2002	Austria	EPrints	Independent
SciRePrints				✓	164	2009	Latvia	EPrints	Institution
Earth sciences									
CEDA Repository		✓			812	2009	UK	EPrints	Institution
Earth-Prints Repository		✓			7,780	2006	Italy	DSpace	Institution
Aquatic Commons			✓		8,072	2007	Belgium	EPrints	Association
Organic Eprints				✓	13,013	2002	Denmark	EPrints	Institution
AgEcon Search			✓		58,007	1995	USA	DSpace	Institution
Open Knowledge Environment of the Caribbean			✓		8	2009	Jamaika	DSpace	Institution
antbase.org				✓	500	1995	USA	In-house	Institution
Social sciences									
Social Science Open Access Repository	✓				21,777	2007	Germany	DSpace	Institution
HAL-SHS	✓				41,416	2003	France	HAL	Institution
eDoc.VifaPol		✓			66,070	2000	Germany	Opus	Institution
Bepress Legal Repository		✓			134,931	2004	USA	In-house	Institution
EduDoc		✓			488	2008	Mexico	In-house	Institution
Fachlicher Dokumentenserver Paedagogik		✓			4,414	2005	Germany	In-house	Institution
Cognitive Sciences ePrint Archive		✓			4,010	1997	UK	EPrints	Independent
African Higher Education Research Online			✓		828	2007	S.Africa	In-house	Institution
Kaleidoscope Open Archive				✓	1,357	2006	France	HAL	Association
PsyDok		✓			2,319	2004	Germany	Opus	Institution
Theory of Psychology Eprint Archive			✓		119	2001	UK	EPrints	Independent
Bibliopsiquis		✓			4,789	2001	Spain	DSpace	Association
Policy Archive			✓		21,935	2008	USA	DSpace	Institution
Archive of European Integration				✓	20,280	2003	USA	EPrints	Institution
Latin American Development Archive			✓		12	2007	USA	EPrints	Institution
Forced Migration Online Digital Library				✓	4,827	2002	UK	Fedora	Institution
Arts and humanities									
Hedatuz			✓		8,133	2002	Spain	EPrints	Institution
ArtXiker				✓	394	2006	France	HAL	Institution
Hprints		✓			116	2008	Denmark	HAL	Institution
ART-Dok		✓			2,551	2007	Germany	Opus	Institution
Propylaeum-DOK		✓			1,536	2007	Germany	Opus	Institution
JIIA Eprints Repository		✓			200	2003	Italy	EPrints	Independent

Note. The topical range is indicated by the check in one of four columns, from left to right: very broad, broad, narrow, very narrow.

exact number of uploaded documents per year. Despite the potential inaccuracies, the data showed that a breakdown into five size categories appeared meaningful. These are briefly presented, together with a discussion of typical chararcteristics.

>100,000 documents. In this category we find seven repositories, including **PubMed Central (PMC)**, **CiteSeer**, **RePEc**, **arXiv**, and **Social Sciences Research Network (SSRN)** as well as the less well known **PhilPapers** and **Bepress Legal Repository**. Four were started in the 1990s, PMC in 2000, Bepress in 2004, and PhilPapers in 2006. All use custom-built software and have broad topics, covering entire or multiple branches of science, and are located in the United States or United Kingdom.

>10,000. Nine repositories fell into this size category. A couple of them are essentially integral parts of larger repositories (**HAL-SHS** and **Munich Personal RePEc** archive). All use third-party software, and all but one were started after 2000. The majority are located in countries other than the United States or United Kingdom, particularly in Germany, and also accept content in languages other than English. Some have more specialized topics.

>1,000. This is the most numerous size category (21), and in many cases, the topics start to become more narrow. These repositories were with two exceptions founded after 2000 and are spread over several countries. Many have received initial funding from organizations such as the European Commission or the United Nations Educational, Scientific and Cultural Organization (UNESCO).

>100. The 13 repositories in this size range are for the most part very narrow in scope, for instance, the Basque language (**ArtXiker**) or digital curation and preservation, a subfield of library science (**ERPAePRINTS**). Quite a few use in-house software.

<100. The five repositories in this size category are largely failed ones, which never reached a critical mass of submissions in order to create a constant flow of new entries. This is very evident in the decline of yearly submissions in later years. It is very likely that numerous similar attempts are still visible on the web, because such repositories might never have registered in either of the two indices. Such repositories would be excellent to study in order to find out why they failed, but this has not been attempted here.

Start Year

The start years of the different repositories were determined by a number of methods, for instance, by checking information on the website, checking upload data of individual manuscripts, and checking the records in the ROAR registry. The age distribution is shown in Table 2, grouped by size category.

TABLE 2. Repositories by start year, grouped in size categories.

Start year	≥100,000	≥10,000	≥1,000	≥100	≥1	Total
1991	1					1
1992	1					1
1993						0
1994						0
1995		1		1		2
1996			1			1
1997	2		1			3
1998						0
1999					1	1
2000	1	1	1	1		4
2001			1	1		2
2002		1	5			6
2003		3		3	1	7
2004	1		2			3
2005			2			2
2006	1	1	3	1		6
2007		1	4	1	1	7
2008		1		3		4
2009		1	1	1	1	4
2010				1	1	2
2011						0
2012						0
Total	6	9	21	13	5	56

Topic Range

One useful way of looking at the repositories is how wide or narrow is their scope. A subjective classification of the repositories into four classes was made: very broad, broad, narrow, and very narrow. The yardstick is roughly how many peer-reviewed journal articles are published per year in that topic, and the range was from almost one third of the whole peer-reviewed literature in the case of **PMC** to a maximum of a few hundred per year for subjects such as ants or information technology (IT) tools in architectural design. *Very broad* would correspond to over 100,000 articles per year (i.e., biomedicine, social science); *broad* to whole scientific disciplines, often with tens of thousands of articles; *narrow* to subfields; and *very narrow* to particular topics. The results were rather even, with 11 very broad, 22 broad, 10 narrow, and 13 very narrow repositories.

Services

The basic service that a repository should provide is permanent storage, a stable web adress for the uploaded manuscripts, and being open to general and web search engines. Most retrievals of the manuscripts would in practice be via search engine hits, for instance, using the titles of the articles. Google Scholar, for instance, has a feature showing openly available, full text copies in a separate column to the left.

Many repositories have features in addition to this basic functionality; examples include:

- Being searchable by special-purpose aggregators (i.e., OAIster), by complying with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

- Possibilty of browsing the repository by subcategories (author, topic, country, year, new submissions, etc.)
- Advanced search facilty within the repository content
- Citation tracking
- Author ranking based on citations
- Reviews of new manuscripts identifying particularly interesting ones
- Other community services such as conference and job announcements

No attempt was made to systematically review such features for each repository. The general impression is that special features could be found especially among the older and larger repositories that use custom-built software. Most of the younger and smaller repositories, which have been built predominantly using open-source repository software, offer the basic set of functionalities.

Organizational Setting

The repositories were classified into three types depending on the organizational setting: those set up by universities, their departments, or other organisations (37), those belonging to international associations (8), and those that have emerged as independent repositories (11) based on initiatives from individual scientists or groups of scientists. Many of the "independent" repositories are nevertheless indirectly supported by universities by the free use of their web servers. The difference is mainly in the history and the level of support by the institution, for instance, paid staff managing the repository.

Some of the more succesful independent repositories have over the years evolved substantially in the organizational sense, from their origins of one individual or a few "entrepreneurs" launching them on their own. **SSRN** has in fact become a corporation with a budget in excess of US$1 million, and an international association has been founded for running **E-LIS**. Some repositories have established hierarchical editorial structures resembling high-quality peer-reviewed journals.

It was also interesting to note that several repositories were started with external initial funding, for instance, from UNESCO, the European Commission, JISC (United Kingdom), U.S. National Science Foundation, Ford Foundation, or Nordbib. Most of these were not particularly successful in attracting a steady flow of new submissions after the intial funding ended.

IT Platform

All the early repositories had to build IT platforms of their own, but later most repositories have been able to use specialized third-party software (i.e., EPrints and DSpace), usually available via open-source licenses. Seventeen repositories use custom-built software, in particular, the largest and most successful ones. EPrints (17 repositories) and DSpace (10 repositories), which initially were designed for institutional repositories, are both very popular platforms for smaller and midsized subject repositories. DLIST (Digital Library of Information Science & Technology) has in fact been implemented as a subject collection in the University of Arizona institutional DSpace repository. HAL (Hyperarticles en ligne) is the French national repository infrastructure, which also has been used to set up a number of subject collections, of which six are included in this analysis. Opus, with four repositories, is a system developed by the University of Stuttgart and is widely used in German-speaking countries, and Fedora (one repository) was originally developed by researchers at Cornell University. SciX was developed within an EU-funded project by Ljubljana University and is used in one repository.

Country of Origin

Not unexpectedly, the United States topped the list with 17 repositories. More surprisingly, Germany tied with the United Kingdom in second place with nine repositories each. The German-based repositories were for the most part rather young, many using the Opus software and also containing a fair proportion of German-language content (in addition, there were two Austrian repositories). France had four and Italy three, with all other countries having at most two.

Repositories by Discipline

There are many alternative ways in which a discussion of the 56 repositories could be organized. The most meaningful seems to be by subject field. The narrative is also partially chronological, in the sense that the fields for which subject repositories were first developed are used here to begin the discussion. Table 1 is structured to follow more or less the same order as the narrative.

Physics and mathematics. Scientists in all disciplines tend to send article manuscripts to a few colleagues to obtain feedback and exchange ideas, but in a few fields such as physics and economics this exchange was systematic even before the Internet and the web, first on paper and later using ftp sites and e-mail list servers. It was thus no coincidence that the first successful repositories emerged in these subject fields. Much has been written about the first successful Eprint archive, **arXiv**, which was started in 1991 by Paul Ginsparg at the Los Alamos National Laboratory. The number of manuscripts uploaded to arXiv has in 2 decades grown linearly to over 800,000, and nowadays the repository includes other disciplines, such as mathematics, nonlinear science, computer science, and quantitative biology. In 2001, Cornell University took over the hosting of the service, and, in 2011, paid university staff had taken over the practical running of the repository. Its current yearly budget is approximately US$400,000, and the university is trying to have institutions with many authors uploading to the repository support the service financially. A collaborative governance structure has also been set up for the repository (Fischman, 2011).

Despite the success of arXiv, several scientists who are dissatisfied with certain aspects of its operation have set up an alternative service called ***viXra***. According to the latter's website, "it has been founded by scientists who find they are unable to submit their articles to arXiv.org because of Cornell University's policy of endorsements and moderation designed to filter out e-prints that they consider inappropriate." The service, founded in 2009, currently houses over 3,000 manuscripts.

Economics and management. The other branch of science that has had a strong preprint culture preceding the web is economics, in which manuscripts have been distributed as working papers published by the university departments and research institutes of the authors. The **SSRN** was started in 1992 under the name Financial Economics Network and was formally incorporated under its current name in 1994. Currently, it has a budget of approximately US$1 million but is largely dependent on voluntary work contributed by over 1,000 scholars worldwide, who act as advisory editors and network directors (Jensen, 2012). The paid staff includes about 15 people in the central office.

SSRN currently stores over 380,000 full papers and abstracts, and, although the majority of papers are available for free, the network includes a number of indexing and abstracting services that are subscription-based. SSRN also includes material from publishers who upload it to SSRN, but this material can be accessed only via pay-per-view.

RePEc (Research papers in economics) was started in 1997, but its origins stretch back to the beginning in 1993, when its predecessor, NetEc, was started. In contrast to SSRN, RePEc is entirely run by volunteers and all services are free. The structure of RePEc differs from most subject repositories in that its backbone consists of a large number (currently over 1,400) of institutional or departmental repositories of working papers and preprints, which are linked together by a common search portal and a number of value-adding services, such as download statistics. Authors who lack a suitable local repository for uploading can use the **Munich Personal RePEc Archive**. There is also an adaptation of the same software and structure in Russian, **Socionet**. In addition, major publishers provide metadata information on published subscription articles.

Both SSRN and RePEc are representative examples of the web portal philosophy, which was very popular in the latter half of the 1990s, in the period when web search engines were not yet fully developed and when readers interested in particular subjects would tend to search for information in discipline-specific hubs.

EconStor is a much more recent initiative (2009) and also has a different type of genesis. Its predecessor was the German National Library for Economics, and it migrated to the DSpace software in 2009. It has a very clear institutional setting and provides repository services to the economics departments of German universities. Its development has been supported partially by the European Commission (the Network of European Economists Online [NEEO] project). EconStor also contains interfaces to RePEc. EconStor is less of a portal where researchers would directly search for publications than a method for a large number of institutions to outsource the technical infrastructure for their repositories.

Computing and information science. It is perhaps no great surprise that researchers in computing and information and library science, given their research areas, have been very active in the creation of subject repositories. **CiteSeer** was started in 1997 by a number of researchers working at the NEC research institute. It was not primarily a repository but rather a search engine for academic content, which harvests the web for openly available literature and also provides citation tracking functions. It can be seen as a forerunner to academic search engines such as Google Scholar. In 2008, it was replaced by **CiteSeer^x**, which has a more scalable software architecture.

E-LIS (e-prints in library and information service), was established in 2003 by an international group of collaborating scholars and has organizationally evolved into an association, with a structure similar to top-notch society-published scholarly journals. The organization includes nearly 50 national editors, who check the metadata of documents uploaded from their countries.

Other repositories in computing and information science have been limited mainly to output from particular countries (the French **Archivesic** and the **Arab Repository for Library and Information Studies**) or to narrow subject fields. Examples of such fields are agent-based computing, cryptology, graph drawing, information systems, and digital curation and preservation. I have personal experience of two such repositories, **Sprouts** for working papers in information systems and **Architektur-Informatik** for IT in architecture. Both were created with great enthusiasm just after the millennium shift, but have never achieved the critical mass of submissions needed for success.

Medicine. The medical field has only few repositories, but one, **PubMed Central (PMC),** is a key resource for the whole OA movement. PMC was developed by the U.S. National Library of Medicine, based on the earlier Entrez search engine for health sciences databases and launched in 2000. In contrast to many of the repositories mentioned earlier, PMC is concentrated on providing OA to manuscript copies of published articles, not working papers or submitted versions. Many publishers have also agreed to deposit exact copies of published papers in PMC, usually with a delay of 12 months. Currently, PMC contains over 2 million OA articles.

Of particular importance for the development of PMC has been the OA mandate of the National Institutes of Health, which has been in place since 2006. This policy requires that articles emanating from NIH-funded research are made openly available in PMC at the latest, 12 months after publication. Because of the NIH's position as the largest public research funding agency in the world, several publishers have lobbied strongly against this mandate in the U.S., but unsuccessfully. With the great popularity of PMC, sister sites have sprung up in other countries (**PMC Canada**, **UK PubMed Central**).

One of the reasons why there have been fewer repositories founded in biomedicine is the absence of a working paper or preprint tradition in this field, and the relatively fast turnaround time from submissions to published articles in medical journals. One attempt to set up a preprint server in the field was **Clinical Medicine NetPrints**, which was sponsored by the BMJ group and Highwire Press. The site contains fewer than 100 manuscripts from 1999 to 2003. The home page of the repository has a very visible warning: "Articles posted on this site have not yet been accepted for publication by a peer-reviewed journal. They are presented here mainly for the benefit of fellow researchers. Casual readers should not act on their findings, and journalists should be wary of reporting them."

A few other repositories in medicine are worth mentioning. **OpenMED@NIC** is hosted by the Bibliographic Informatics Division of the National Informatics Centre (India) and is intended for both preprints and green copies of accepted manuscripts. It seems to be used mainly by Indian authors. **Dryad** offers authors of published medical articles the possibility to upload data sets related to their journal articles to the repository.

Philosophy. Scholars in philosophy seem to have embraced the idea of subject repositories eagerly. **PhilPapers** also includes types of materials other than just OA copies of articles and has many characteristics of a "one-stop shop" portal for scholars in this domain. As with many other subject repositories, it has been developed (since 2006) by a couple of entrepreneurial scientists, although it has received sponsorship from JISC in the United Kingdom. **Sammelpunkt Elektronisch Archivierte Theori** is a small repository with a wide variety of subjects, although the majority of papers are in some field of philosophy. It contains papers in both German and English.

Two repositories with narrower subject areas are worth mentioning. The **Philosophy of Science** archive was set up in the year 2000 by scholars at the University of Pittsburgh, inspired by the success of arXiv. Like its role model, it concentrates on preprints. **SciRePrints** is a small repository hosted by the University of Latvia, focusing on papers discussing the relationship between science and religion. The home page contains an interesting passage: "Notably, scientific articles not accepted for publishing in other sources due to religious or mystical presuppositions are welcome here, provided that they comply with the academic standards and use scholarly methods and language."

Earth sciences. There are several repositories within the general area titled earth sciences, including two broad ones dealing with earth and atmospheric sciences (**CEDA**, **Earth Prints**) and more specialized ones dealing with topics like marine environment research (**Aquatic Commons**), organic food and farming (**Organic Eprints),** and research related to the Carribean region. A successful repository with a critical mass of material is **AgEcon Search**, hosted by the University of Minnesota, which specializes in agricultural and applied economics (Kelly & Letnes, 2006). Another narrowly focused repository is **Antsbase**, a collaboration between the American Museum of Natural History and The Ohio State University.

Social sciences. In addition to economics and management, there are several repositories in the social sciences but only three with a significant critical mass of papers exceeding 10,000 manuscripts (**Social Science Open Access Repository**, the German **eDOC.VifaPol** for administrative and political science, and **HAL-SHS**). The last of these is one of the overlay structures providing subject views, in this case, social sciences, into the national French HAL repository. In addition, there are several more specialized repositories, for instance, five in education and pedagogy and three in psychology and psychiatry. An interesting repository combining a subject and a regional aspect is **African Higher Education Research Online**.

In history research, several repositories listed in ROAR and DOAR had to be discarded because they contained digitized documents only, but there were interesting repositories dealing with, for instance, European integration and Latin American development. A repository that seems to be functioning very well is the **Forced Migration Online Digital Library**, which aims to collect a multitude of information resources related to refugees and forced migration (Cave, Loughna, & Pilbeam, 2008). The website is of a high quality, and the repository also aims to reach out to policy makers, the broader public, and teachers. The website also provides a facility to donate money for the maintenance of the site.

Arts and humanities. Several highly specialized repositories in the arts and humanities exist, but no broader ones. It is important to note that scholars in these fields tend to publish more in monographs or book chapters and that peer-reviewed journal publishing is less common compared with the scientific-technical-medical fields.

Two repositories deal with different aspects of Basque culture, one broader (**Hedatuz**) and one concentrated on the language (**ArtXiker**). Both accept inputs in several languages (Basque, French, Spanish, English). The latter, like **hprints.org** (the free Nordic arts and humanities and social sciences e-print repository), uses the French national repository infrastructure HAL. Hprints.org was started with Nordic funding but never really took off. Other repositories with narrow domains exist for classical studies (**Propylaeum-DOK**), art history (**ART-Dok**), and archeology (**JIIA**).

Discussion

The evolution of subject repositories must be seen in context against the general development of the Internet and the development of the OA movement. Several of the leading repositories had been already developed in the mid-1990s, when most scholarly journals were still distributed only as paper copies and when creating portals to web information and link lists was more important than today. Since then, almost all major publishers have started parallel electronic publishing of subscription journals, and a rapidly increasing number of universities and research institutes have launched institutional repositories of their own, which compete with subject repositories for the same papers.

Only the largest subject repositories contribute significantly to the overall volume of green OA copies. A recent study showed that 43% of self-archived manuscript copies are located in subject repositories (Björk et al., 2010), and 94% of these were located in either arXiv or PMC (Björk et al., 2013). Other repositories may play an important role in their niche areas, but there are so many blank research areas without a relevant subject repository that the overall effect is very small. Areas that in the context of this study lack significant repositories include chemistry and engineering.

Determining whether a particular repository is successful is a difficult and subjective task. On a theoretical level, methods for measuring this for any type of repository have been discussed by Thibodeau (2007), who states that success is measured by "how well it covers the universe of assets it should or might hold." Adamick and Reznik-Zellen (2010b) also discuss this issue. The success or critical mass of a repository should ideally be judged by comparing the actual uploaded content to the potential uploadable literature in the topical range. In practice, it would be very difficult, however, to determine the potential article volumes for many of the repositories, unless their topics coincide exactly with disciplines defined in Web of Science (WoS) or Scopus. It would be equally difficult to determine the exact numbers of WoS- or Scopus-indexed articles among the documents uploaded to the repositories, because the repository may contain a wide variety of material other than copies of peer-reviewed articles. A pragmatic measure that is easier to use is the *trend* in the number of uploads, which also has been discussed by Carr and Brody (2007). Authors in a field will soon lose faith in a repository that has not achieved a critical mass of articles and will stop uploading new documents. For some of the smaller repositories, which have been mentioned as failures, this criterion has been used.

It is interesting to note the wide variety of organizational structures that has emerged around subject repositories. The most common history is that of a single or a handful of "entrepreneur" scholars who have created the repository as a more or less personal project. Usually, their institution has allowed the use of the university website. In some cases, the development has later led to a corporate structure with employed staff (SSRN), and in others to the emergence of complex networked voluntary work structures (RePEc). Repositories that were started by institutions based on a strategic decision by top management (i.e., PMC, HAL) are rare. Armbruster and Romary (2009) note that "the future of subject-based repositories depends on whether they develop a sustainable business model with independent income."

The results of this study can be compared with results from Adamick and Reznik-Zellen (2010b), who studied the 10 largest repositories also included here. The following observations can be made:

- The geographical spread of the home countries of the repositories becomes much more diverse (outside the United States) as we go outside the "big 10." Several midsized and small repositories also welcome uploads in languages other than English.
- Midsized and smaller repositories tend to have more narrow topical ranges than the large ones.
- There are several smaller repositories in niche areas in the social sciences, arts, and humanities.
- The use of open-source software is dominant outside the big 10.
- Many failed repositories are included. The business model in which a repository was started with time-limited funding from an outside grant seems not to have been particularly fruitful.

Conclusions

Despite the availability of open-source repository software (i.e., EPrints, DSpace), which technically has lowered the initial effort needed to start a repository, it seems that the potential for launching new successful repositories has diminished. It is currently from a managerial viewpoint much easier to launch an institutional repository, which is usually operated by dedicated university library staff, than a subject repository, which may lack initial funding and may require an international network of collaborators to get started. An institutional repository is the natural locus of PhD and lower theses and already existing working paper and publication series, and it can be backed by a mandate from the university making it obligatory for the institution's researchers to upload green copies of their journal publications. Institutional repositories are also natural extensions to the current research information systems that almost all universities now use to keep track of their publication output, and it is very "trendy" to start institutional repositories. According to Björk et al. (2013), 82% of the world's 148 most productive research institutions have an institutional repository offering a natural place to self-archive for 85% of the articles produced in these institutions. Subject repositories, on the other hand, must rely mainly on word-of-mouth within their communities and on reaching the necessary critical mass early in order to get underway.

Another development of importance is the legal boundary conditions for self-archiving. Although a majority of publishers allow the open self-archiving of the final, approved manuscript version, this is usually allowed only for author web pages and institutional repositories; usually, subject repositories are excluded. In a study of the copyright rules of the 100 largest scholarly publishers, immediate self-archival of the accepted version was allowed for 61% of all published articles in institutional repositories but for only 21% in subject repositories (Laakso, 2013). Such detailed copyright rules started to become common in 2003–2004, and apparently publishers believe that subject repositories are a greater threat to their business. The only exception is PMC, which, with the bargaining power of the NIH as research funder, has been able to negotiate special conditions with several major publishers.

The few large, successful subject repositories are likely to continue to thrive, because they have become a part of the publishing behavior of academics in their fields. In particular, two factors have contributed strongly to the emergence of successful subject repositories in a limited number of areas. The first is the existence of a strong working paper or preprint culture in a research field in place prior to the Internet (as was the case for arXiv, SSRN, and REPEC). The second is a mandate from a dominant research funder to upload copies to a prescribed subject repository (PubMed Central). If new mandates from strong research funders would emerge (for instance, through stakeholders such as the U.S. government or the European Commission), these might help support the growth of existing or new subject repositories, but the trend in such mandates seems to be to promote self-archiving in institutional repositories as well, which is not the case in the NIH mandate.

All in all, it seems that the strongest growth period for subject repositories is over. The growth in green OA literature available via subject repositories currently consists mainly of the internal growth of the few really large repositories (PMC and arXiv, in particular) rather than the emergence of new repositories.

Acknowledgments

I thank the two anonymous reviewers for very constructive comments, which helped to improve the manuscript.

References

Adamick, J., & Reznik-Zellen, R. (2010a). Representation and recognition of subject repositories. D-Lib Magazine, 16, doi:10.1045/september2010-adamick

Adamick, J., & Reznik-Zellen, R. (2010b). Trends in large-scale subject repositories. D-Lib Magazine, 16, doi:10.1045/november2010-adamick

Armbruster, C., & Romary, L. (2009). Comparing repository types: Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. Working paper, November 23, 2009. Retrieved from: http://ssrn.com/abstract=1506905

Björk, B.-C., & Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. BMC Medicine, 10:73, doi:10.1186/1741-7015-10-73

Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Gudnasson, G. (2010). Open access to the scientific journal literature: Situation 2009. PLoS One, 23.6.2010. doi:10.1371/journal.pone.0011273

Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2013). Anatomy of green open access. Journal of the American Society for Information Science and Technology (in press).

Carr, T., & Brody, T. (2007). Size isn't everything—Sustainable repositories as evidenced by sustainable deposit profiles. D-Lib Magazine, 13, http://www.dlib.org/dlib/july07/carr/07carr.html

Cave, M., Loughna, S., & Pilbeam, J. (2008). Open access repository system for forced migration online. Association of Librarians and Information Professionals Quarterly, 3(4).

Darby, R., Jones, C., Gilbert, L., & Lambert, S. (2008). Increasing the productivity of interactions between subject and institutional repositories. New Review of Information Networking, 14, 117–135.

DeRobbio, A., & Katzmayr, M. (2009). The management of an international open access repository: the case of E-LIS. GMS Medizin—Bibliothek Information, 9(1), http://www.egms.de/static/pdf/journals/mbi/2009-9/mbi000137.pdf

Fischman, J. (2011) The first free research-sharing site, arXiv, turns 20 with an uncertain future, Chronicle of Higher Education, 10.8.2011. Retrieved from: http://chronicle.com/blogs/wiredcampus/the-first-free-research-sharing-site-arxiv-turns-20/32778?sid=wc&utm_source=wc&utm_medium=en

Ginsparg, P. (2004). Scholarly information architecture, 1989–2015. Data Science Journal, 3, 29-41. Retrieved from: https://www.jstage.jst.go.jp/article/dsj/3/0/3_0_29/_pdf

Jensen, M. (2012) About SSRN, 2.2.2012. Retrieved from: http://www.ssrn.com/update/general/mjensen-20th.html

Kelly, J., & Letnes, L. (2006). Managing the grey literature of a discipline through collaboration: AgEcon search. Resource Sharing & Information Networks, 18, 157–166.

Kim, J. (2010). Faculty self-archiving: Motivations and barriers. Journal of the American Society for Information Science and Technology, 61, 1909–1922.

Kleinman, M. (2011). Faculty self-archiving attitudes and behavior at research universities—A literature review, PhD term paper, University of Michigan. Retrieved from: http://mollykleinman.com/wp-content/uploads/2012/02/Kleinman-self-archiving-literature-review-web.pdf

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, 1306–1320.

Laakso, M. (2013). Journal publisher self-archiving policies and the potential for growth in open access. Working paper, Hanken School of Economics, Helsinki, Finland.

Laakso, M., & Björk, B.-C. (2012). Anatomy of open access publishing—A study of longitudinal development and internal structure. BMC Medicine, 10, 124. doi:10.1186/1741-7015-10-124.

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, 195–206.

Parinov, S., & Krichel, T. (2004). RePEc and Socionet as partners in a changing digital library environment, 1997 to 2004 and beyond. In: Russian Conference on Digital Libraries, Puschchino, Russia. Retrieved from: http://eprints.rclis.org/1830/

Suber, P. (2012). Open access. Boston: MIT Press. http://cyber.law.harvard.edu/hoap/Open_Access_(the_book)

Swan, A. (2010). The open access citation advantage: Studies and results to date. Key Perspectives Report. Retrieved from: http://eprints.ecs.soton.ac.uk/18516

Thibodeau, K. (2007). If you build it, will it fly? Criteria for success in a digital repository. Texas Digital Library, 8, http://journals.tdl.org/jodi/index.php/jodi/article/view/197/174

Wagner, B. (2010). Open access citation advantage: An annotated bibliography. Issues in Science and Technology Librarianship. doi:10.5062/F4Q81B0W

Xia, J. (2008). A comparison of subject and institutional repositories in self-archiving practices. Journal of Academic Librarianship, 34, 489–495.