



HAL
open science

Degrees of Openness

Hélène Prost, Joachim Schöpfel

► **To cite this version:**

Hélène Prost, Joachim Schöpfel. Degrees of Openness: Access Restrictions in Institutional Repositories. D-Lib Magazine, 2014, 20 (7/8), 15 p. 10.1045/july2014-prost . sic_01077190

HAL Id: sic_01077190

https://archivesic.ccsd.cnrs.fr/sic_01077190

Submitted on 10 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Degrees of Openness. Access Restrictions in Institutional Repositories

Hélène Prost

Joachim Schöpfel (corresponding author)

Bio notes

Hélène Prost is information professional at the Institute of Scientific and Technical Information (CNRS) and associate member of the GERiiCO research laboratory (University of Lille 3). She is interested in empirical library and information sciences and statistical data analysis. She participates in research projects on evaluation of collections, document delivery, usage analysis, grey literature and open access, and she is author of several publications.

Joachim Schöpfel is Lecturer of Library and Information Sciences at the University of Lille 3 (France), Director of the French Digitization Centre for PhD theses (ANRT) and member of the GERiiCO research laboratory. He was Manager of the INIST (CNRS) scientific library from 1999 to 2008. He teaches Library Marketing, Auditing, Intellectual Property and Information Science. His research interests are scientific information and communication, especially Open Access and Grey Literature.

Contact

Joachim Schöpfel

Charles de Gaulle University Lille 3

GERiiCO laboratory - ANRT

BP 60149

59653 Villeneuve d'Ascq Cedex

France

joachim.schopfel@univ-lille3.fr

Tel ++33 (0) 320 41 61 06

Abstract

Institutional repositories, green road and backbone of the open access movement, contain a growing number of items that are metadata without full text, metadata with full text only for authorized users, and items that are under embargo or that are restricted to on-campus access only. This paper provides a short overview of relevant literature and presents empirical results from a survey of 25 institutional repositories that contain more than 2m items. The intention is to evaluate their degree of openness with specific attention to different categories of documents (journal articles, books and book chapters, conference communications, electronic theses and dissertations, reports, working papers) and thus to contribute to a better understanding of their features and dynamics. We address the underlying question of whether this lack of openness is temporary due to the transition from traditional scientific communication to open access infrastructures and services, or here to stay, as a basic feature of the new and complex cohabitation of institutional repositories and commercial publishing.

Introduction

Open archives are less open than they should be. In particular, institutional repositories contain a growing number of metadata without full text or with full text only for authorized users. This is not in line with the underlying principles of the open access (OA) movement that defines open access “as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community”¹ and that requires freely available scientific literature provided by open archives (*green road*) and OA journals (*gold road*). Freely available means “free of charge, and free of most copyright and licensing restrictions” (Suber 2012, p. 4).

Twenty years after the arrival of the first open repository called *arXiv*² and ten years after the Berlin declaration on open access, the situation is not really satisfying. On the one hand, the open access movement has undergone “dramatic growth”³, with nearly 10,000 OA journals, about 3,500 repositories and several millions of freely available documents on the Internet. On the other hand, the repositories, *green road* and genuine backbone of the OA movement, are less open than expected. They contain many scientific documents that were not available previously on the Internet, but some items are under embargo or restricted to on campus access, and for other items there is only metadata, without links to the full text.

Is the glass half empty or half full? Is this lack of openness temporary due to the transition from traditional scientific communication to OA infrastructures and services? Or is it here to stay? Are embargoed documents the price to pay for the accelerated development of institutional repositories?⁴ Although the answers to these questions will be political, they but should take into account and reflect reality. So, what can we conclude about content with restricted access? Are repositories more open for some document types than for others? This paper provides a short overview of relevant literature and presents empirical results from a survey of 25 institutional repositories that contain more than two million items. The intention is to evaluate their degree of openness with specific attention to different categories of documents and thus to contribute to a better understanding of their features and dynamics.

Literature overview

Institutional repositories (IR) have been defined as “tools (...) for collecting, storing and disseminating scholarly outputs within and without the institution” (Jain, 2011) and as “a set of services (...) for the management and dissemination of digital materials created by the institution and its community members (based on) organisational commitment to the stewardship of these digital materials” (Lynch, 2003). They serve “the interests of faculty – researchers and teachers - by collecting their intellectual outputs for long-term access, preservation and management” (Carr et al., 2008). As the reason for setting up a repository “carries implications for the content, design and funding of a repository, (...) the institution

¹ Berlin declaration on Open Access to Scientific Knowledge of 22 October 2003, available at

<http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>

² <http://arxiv.org/> with more than 900,000 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics.

³ <http://poeticeconomics.blogspot.fr/2006/08/dramatic-growth-of-open-access-series.html>

⁴ See for instance Stevan Harnad’s comments on recent UK decisions on open access: “The mandate must uncouple the date of deposit from the date the deposit is made OA, requiring immediate deposit, with no exemptions or exceptions. How long an OA embargo it allows is a separate matter, but on no account must date of deposit be allowed to be contingent on publisher OA embargoes (...) In reality, embargo is but one pb, and not the most important.” <http://openaccess.eprints.org/index.php?/archives/1103-HEFCEREF-Adopts-Optimal-Complement-to-RCUK-OA-Mandate.html>

needs to be clear about the implications of different roles for a repository, while being prepared to change or add roles as the scholarly communication environment develops” (Friend, 2011).

However, they are not homogeneous. There is not one model but multiple options and realizations, and they show different policies, procedures, functionalities, services and metadata, with different business models and funding strategies (Swan & Awre, 2006). Also, their content may include more than current output from faculty. Smith (2008) details a “wide variety of materials in digital form, such as research journal articles, preprints and post prints, digital versions of theses and dissertations, and administrative documents, course notes, or learning objects.” Other repositories include datasets, multimedia or cultural and scientific heritage.

The exact number of open archives is unknown. The statistics of the main international directories⁵ vary between 2,600 and 3,600 sites but the real number is probably higher. OpenDOAR counts 2,616 repositories of which 2,163 are listed as institutional (83%). The ROAR directory contains 2,388 institutional or multi-institutional repositories, that is 66% of all sites (3,621). In spite of different figures, there is no doubt that these sites represent the most important part of the so-called *green road* to open access.

This does not mean that their content is 100% open and freely available and accessible. Some contain bibliographic references without full text. Others protect access rights. Only a small number of them clearly define a content policy. The OpenDOAR directory warns the user of its repository content search engine that “full texts are not available for most results” but does not provide any statistics. Operated by the Bielefeld University Library, the search engine BASE provides more than 50 million documents via the "Open Archives Initiative Protocol for Metadata Harvesting" (OAI-PMH). According to BASE the full text is available for about 75% of the indexed documents. Yet, a quick browse on different document types shows something very different:

Document type	Total items	Open access	Unknown access
Articles	12418039	15,90%	84,10%
Books	1914181	13,88%	86,12%
Theses	2717008	5,10%	94,90%
Datasets	1110490	0,01%	99,99%

Table 1: Access status of items retrieved with BASE (February 2014)

Only a small percentage of retrieved items are clearly open access. For most of the repository content, the BASE search engine indicates an “unknown access” status. Of course, unknown does not necessarily mean restricted or no access. Nevertheless, as our own research in the field of electronic theses and dissertations shows (Schöpfel & Prost 2013a), a significant part of the “unknown access” content is indeed not freely available but under embargo, available only for authorized users and/or on the academic campus or via the institutional intranet, and some of them are available only on a publisher’s platform. For PhD theses, our non-representative sample of institutional repositories produced the following

⁵ OpenDOAR <http://www.opendoar.org>, ROAR <http://roar.eprints.org/>, BASE. The following statistics were downloaded on March 30, 2014.

data: out of 26% deposited PhD theses with limited access, 17% are embargoed for six months to two years or longer and 9% can only be accessed on-campus (Schöpfel & Prost 2013b).

A recent paper from Spain provides interesting figures about openness of the institutional repository of the Spanish National Research Council CSIC, showing significant differences between collections of research institutes and document types, together with correlations between openness and full text download statistics (Bernal 2013). The following study tries to take a closer look at these figures and to compare them with other repositories.

Methodology

The empirical data in our study are from a sample of 25 institutional repositories. All repositories were selected using the repository search tool OpenDOAR, the authoritative directory of academic open access repositories. The following search criteria were applied:

- Repository type: Institutional
- Content type: PhD theses and articles (at least)
- Size: 10,000+ items

The search was conducted by region (Europe, Asia, Africa, Australasia, North America, South America/Central America/Caribbean), and only those repositories that are operational (i.e. recently updated), that contain different document types including non commercial literature (theses, reports etc.), that allow for filtering by document type and access options (full-text vs. restricted/no access to full-text) as a browse and/or search functionality and that indicate the exact number of results (retrieved items) were selected.

Secondly, we conducted a detailed search and/or browsed on each site for specific document types: articles, books and book chapters, conference proceedings and communications, reports, PhD theses, and working papers (unpublished). We also looked for patents and datasets but did not include them in the global analysis. For each document type, we distinguished the items with free and non-restricted access to the full text (open access) from those with restricted access (embargo, intranet, authorized users, etc.) or without full text (reference only). Whenever possible, we also made this distinction for the entire repository content.

The repositories were selected in February 2014. The analyses of each site were conducted in February and March 2014.

Results

Size and openness of the repositories

The selected repositories (IR) compliant with the criteria outlined above are listed in the appendix. For our study, we did not evaluate the whole content of each IR but limited the analysis to six document categories (working papers, theses, reports, articles, communications, books/book chapters). The total number of items in our study is 2,086,622.

The median size of the sample repositories is 26,683 documents, ranging from 1,199 (Amherst) to 775,561 (HAL). Again, this is *not* the total size but the sum of the selected and evaluated document types, excluding for example courseware, images or Master dissertations; thus, the true size is higher.

The median degree of openness of all repositories is 0.38 which means that only close to 2/5 of all items provide open access to the full text. The individual repositories range from 0.04 (only 4% of items have full text) to nearly 1.00 (except for a few items, all deposits have freely available full text). Figure 1 combines repository size and degree of openness, ranking the IR from most open (left side) to nearly closed (right side). The size of the dot corresponds to the number of items of the repository.

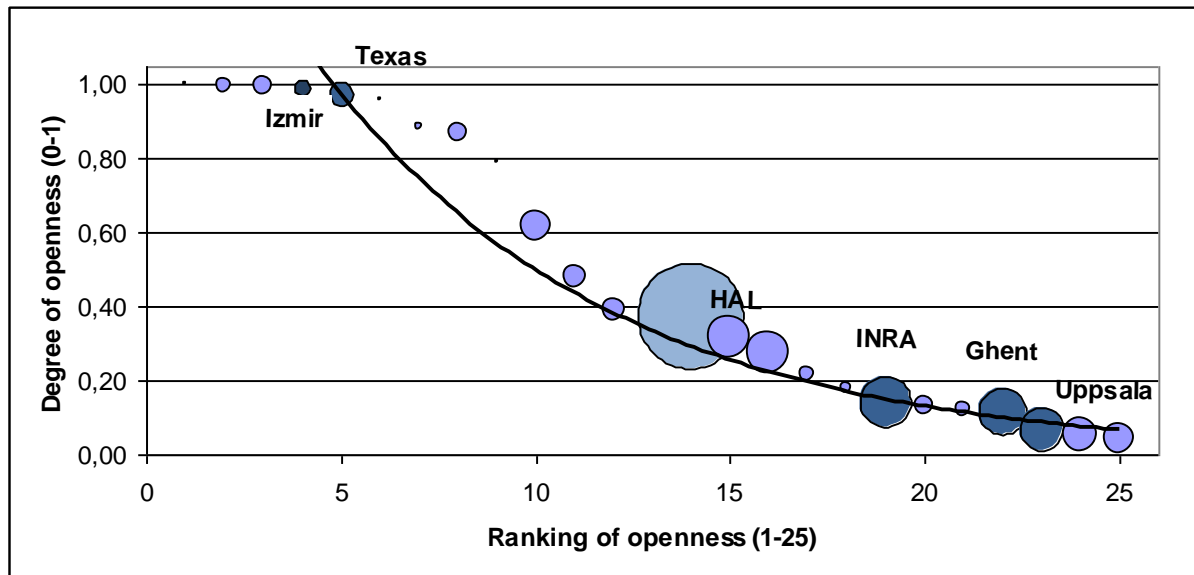


Figure 1: Openness and size (dot) of repositories, with regression line (exponential tendency)

Figure 1 shows that the IRs in our sample are not similar, with a significant variation between smaller but open repositories (such as Izmir and North Texas Denton) and larger sites that are less open (such as HAL, ProdINRA, Ghent or Uppsala).

Openness per document type

More than half of the documents in our sample are journal articles. Together with the conference communications, they represent more than $\frac{3}{4}$ of the entire content. The six document categories in the sample were distributed in the manner shown in Figure 2.

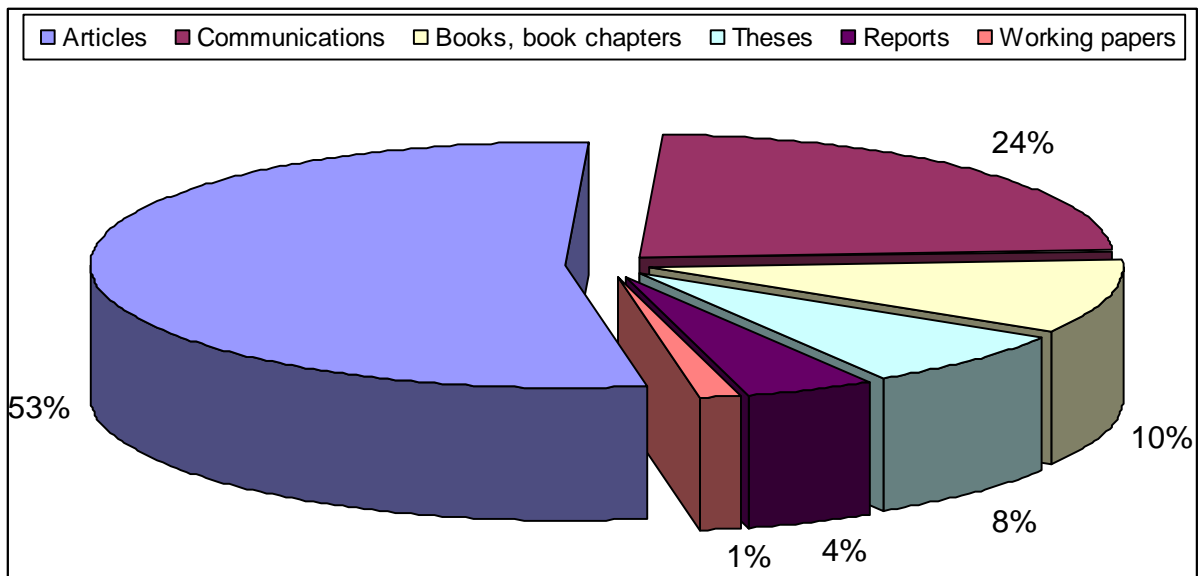


Figure 2: Document types in the institutional repositories (N=2,086,622)

Compared to articles and communications, the other document types are less important. Books and book chapters are represented at 10%, followed by PhD theses (8%), reports (4%) and working papers (1%). The evaluation of their degree of openness – the part of the items freely available on Internet – offers specific values for each document type (figure 3).

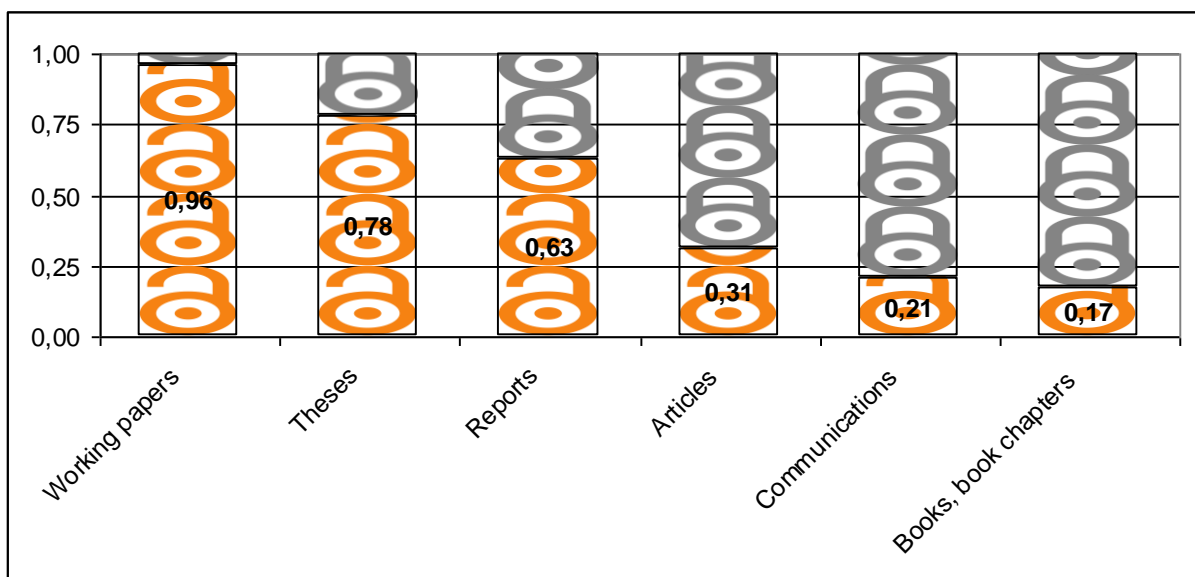


Figure 3: Degree of openness per document type (N=2,086,622)

The overall degree of openness of working papers is 0.96, which means that in the entire sample all but 4% of the working papers are freely accessible' followed by PhD theses (0.76)

and reports (0.63). Significantly less open are journal articles (0.31), communications (0.21) and books/book chapters (0.17). In other words, articles are half as open as reports, and PhD theses are over four times more open than books or book chapters.

The median degree of openness per repository confirms the overall statistics. The median is high for working papers (0.98) and theses (0.92), medium for reports (0.63), and low for articles (0.38), communications (0.29) and books (0.13). The variance of openness (dispersion from average) is relatively low for working papers and theses, while the other categories are more dispersed.

However, we must be careful with interpretation because all of the IRs have articles and theses, most have reports, communications and books, but only half of the IRs have working papers which reduce the variance.

A last observation: the number of items and their openness are inversely correlated, in that the more important categories (articles, communication and books/book chapters) are less open than the less important ones (table 2).

	Openness	Nb of items
Working papers	0.96	31,180
Theses	0.78	156,546
Reports	0.63	91,069
Articles	0.31	1,116,199
Communications	0.21	490,442
Books, book chapters	0.17	201,186

Table 2: Openness and number of items per document type (N=2,086,622)

Is lack of openness the price to pay for large numbers of items? Again, we must be careful with interpretation as there may not be any causal relationship. So far, articles and communications remain the most important part of scientific communication and both are more or less controlled by commercial publishing, with a higher degree of copyright protection. We'll come back to this point in the discussion.

Profiles of repositories

The degree of openness of repositories is mainly influenced by the percentage of freely available articles, as they are the most important part of the repositories' content. Content structure and openness are closely related. Repositories with a high percentage of articles with full text most often have a higher degree of openness, with a Pearson coefficient near to 1.0.

A scattergram with the percentage of open ETD and articles reveals three clusters of institutional repositories (figure 4).

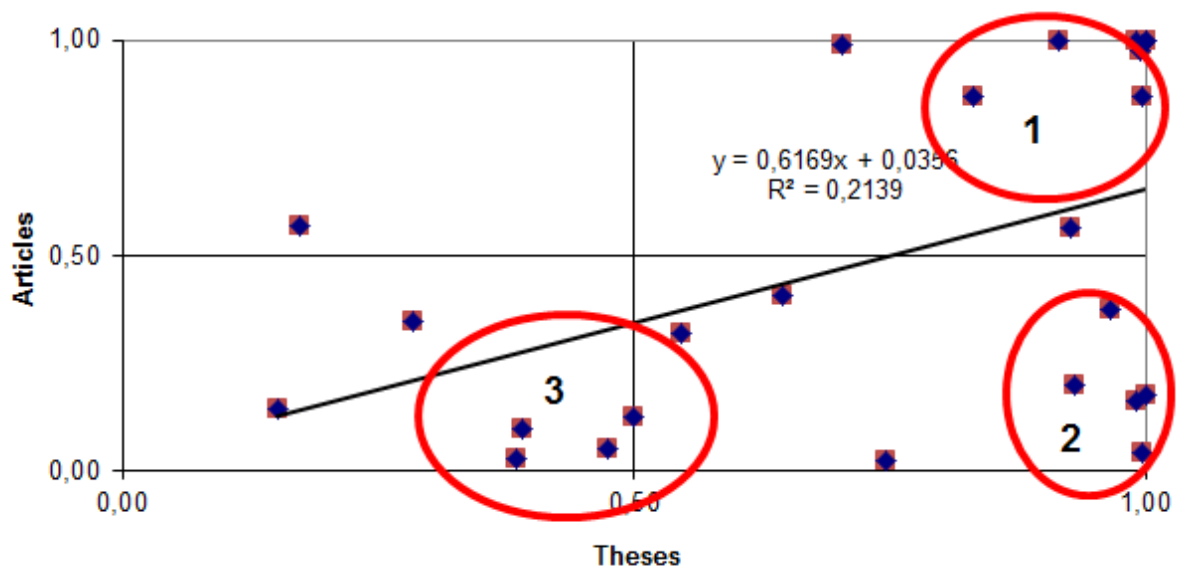


Figure 4: Openness of theses and articles (all repositories)

Based on the median and quartiles, we can distinguish three clusters of repositories:

(1) Articles and theses with a high degree of openness (upper right side): among these are the repositories of Milano, Geneva, Izmir and Chiba.

(2) Theses with a high degree of openness, articles with medium or low openness (lower right side): the repositories of CSIC, CNRS (HAL), Swinburne or the Royal Melbourne Institute of Technology are in this category.

(3) Repositories with a medium or low degree of openness for articles and a medium degree of openness for theses: included here are Torino, Ghent and Uppsala (in the middle of the lower side).

If we take into account all documents, not only articles and theses, the landscape of repositories becomes more complex and differentiated. Yet, we can identify four different groups with regards to openness:

Open repositories: ten sites are more open than the others, for all or most document types (median 0.97, range 0.62-1.00). Examples: Milano, Geneva, Izmir, Chiba, CSIC, Western Kentucky.

Closed repositories: six sites have lower degrees of openness for all or most document types (median 0.13, ranging 0.07-0.39). Examples: Torino, Ghent, Uppsala, National Taiwan University, Chalmers (Göteborg).

Mixed repositories: six sites with different degrees of openness for different types of documents (median 0.17, ranging 0.05-0.48). Examples: Brisbane, Swinburne, Monash University Melbourne.

Grey repositories: three sites relatively open for grey literature and relatively closed for articles and books (median 0.27, ranging 0.18-0.37). Examples: HAL, Hong Kong, Singapore Management University.

The characteristics of these four groups was determined empirically and may only reflect the particularities of the sample and the selected document types. Nonetheless, they may illustrate the emerging and heterogeneous landscape of institutional repositories and reveal different evolutions, policies and environments.

Other results

Some institutional repositories contain datasets and patents. We identified nearly 70,000 items in our sample; 60,219 datasets and 8,982 patents. While only 3% of the datasets were freely available, patents are disseminated with a degree of openness of 0.61, which means that nearly 2/3 of the patents are freely accessible in these repositories. This is surprising for two reasons: the global tendency for open data and the often high protection of patents.

Often the real nature of access restriction remains uncertain. Are the documents under embargo and will they be released and openly accessible in the future? Are they restricted to on-campus access only or is it both of these? And what about missing full text, records without documents? From our results we can only make a cautious guess, that embargo periods represent a small part of access restrictions (in our sample only 2%) and that most of the lack of openness is caused by on-campus only access and by the deposit of metadata without a corresponding document.

Finally, our study was not designed as a longitudinal survey to detect developments and tendencies over three years or more. Yet, our data allow for some anecdotal evidence. The degree of openness of the Spanish CSIC repository declined between 2007 and 2014 from 0.99 to 0.56, which confirms other observations that with time, the repositories not only contain more and more items and metadata but also become, at least partially, less open.

Discussion

Sampling, searching and browsing

Our intention was to focus on large institutional repositories with 10,000 items or more and to select a random sample that met the criteria described above, from all geographic regions. However, this task was more difficult than expected. In some regions, such as South America or Africa, there are few repositories with more than 10,000 items that also have a rich content, not just theses and dissertations. In North America, but also in Africa, many repositories lack advanced functionalities of filtering and browsing so that it is impossible to identify OA items and specific document types.

In other repositories, the meaning of open access or full text is ambiguous. We found different options, like open access, full text, PDF, open access via publisher. The last option is a very special interpretation of open access because the link to the publisher's server is generally restricted to authorized users and is not open at all. Two examples from Australia: [espace@curtin](mailto:espace@curtin.edu.au)⁶, Curtin University's institutional research repository at Perth (Australia), does not allow for filtering open access items; the search results explain that the "file (is) restricted" or provide an alternative location linking through DOI to the publisher's platform.

⁶ <http://espace.library.curtin.edu.au>

The University of New England repository e-publication@UNE⁷ provides links to servers or publishers' platforms with restricted access or to the local OPAC "where you can borrow or buy the book". Sometimes one must register for searching in the repository. Is this still open access?

For our study, these factors may introduce a bias. Because of the selection criteria, the sample excludes not only small repositories but also repositories without advanced and rich functionalities; that is, those with basic indexes, search and browsing options. Only a survey of the hosting institutions could produce empirical evidence for these repositories.

From documents to items

Our study reveals another different kind of problem. Our intention was to measure the degree of openness for institutional repositories and also for some main document types. This was not always possible for two reasons.

The first reason is that in a large number of repositories, it is simply not possible to browse or search for specific document types. All deposits are considered "items" without the traditional library distinctions of articles, books, dissertations and so on. This is all the more regrettable especially as these distinctions are helpful not only for librarians but also for readers. When the complete metadata sets can be visualized, we realize that there are two situations. Sometimes the metadata describe the document type but they cannot be searched or browsed. In the other case, repositories simply have not indexed the document type at all.

The second difficulty is the complete lack of standardization or even harmonization. The typologies are more or less detailed, depending on local needs and habits. Some repositories distinguish between three different types of reports or five different theses and dissertations while others do not. Sometimes even the sub-collections of the same institutional repository - institutes, departments, schools and so on – describe their document types in different ways.

Retrieving datasets is particularly difficult. We were looking for research results, raw data or small data, unpublished material that could be reused for verification, replication, data-mining or meta-analysis. But only some of the repositories correctly index them as data. Others split them up into categories such as sound, speech, survey, still image, script and so on.

Regarding lack of differentiation vs. too much differentiation... Is it still necessary to distinguish between document types? Is it enough to provide access to "information"? Is it not better to transform documents into "items"? We don't think so, because the typology of documents contains valuable information for the reader, about quality and labelling and so on, and because a minimum of standardization is necessary for the interoperability of all these sites.

The case of grey literature

The analysis of the different degrees of openness in institutional repositories reveals differences not only between repositories but also between document types (see figure 3 and table 2 above). Some of these documents, in particular theses, working papers and reports, are grey literature, defined as "not controlled by commercial publishers" (Schöpfel & Farace 2010).

⁷ <https://e-publications.une.edu.au>

Compared to articles published in journals, books and book chapters, these categories are generally more available via open access. In sixteen repositories, their degree of openness is higher than for articles and books, and in six others, they are at the same level. Institutional repositories seem to facilitate the dissemination of grey literature in open access, above all, working papers and also theses and reports.

However, some repositories display rather low degrees of openness for theses and/or reports, with only one or two items out of five available in full text. This lack of openness cannot be explained by assignment of rights to publishers. The reasons are different and especially for theses and dissertations, the decision to embargo or deposit metadata without the full text can be explained by lack of awareness, intellectual property concerns and fear of plagiarism, legitimate interests, expected exploitation (publishing) and trade secrets (Schöpfel & Prost 2013a). The reasons for restricted access to technical and scientific reports which are often institutional products may be different and more related to confidential or sensitive content, trade secrets and so on.

Communications, contributions to scientific conferences, workshops and seminars, are somewhere between both categories. Their degree of openness is often lower than theses, reports and working papers and higher than articles and books, yet closer to the last categories. The explanation here may be very wide-ranging copyright protection, with some types of communications published through commercial channels, as special issues or parts of journals along with articles or in book series, while others are disseminated as grey literature by institutions, learned societies, non-commercial publishing houses, etc.

In the future, institutional repositories should be very careful, and aware of the amount of interest there is in these types of documents. Often, articles and books are available through other channels and in other versions, in particular on the publishers' servers. On the contrary, grey literature, because of its generally strong institutional character, is often available on a limited number of servers or only on one platform, the institutional repository. Open access to these items should therefore be provided whenever possible, without embargos or other access restrictions.

Conclusion

The basic idea of open access may be simple, to cite Peter Suber (2012), but the reality of the open access movement is composite and multifaceted. Contrary to expectations, open does not always guarantee access to the documents, and too often institutional repositories which became the main vector of the OA movement give priority to large numbers of records over high degrees of openness.

Two strategies contribute to this situation. On the one hand, institutions have seized the opportunity of institutional repositories to gain control of their own scientific output. Large and exhaustive repositories allow for scientometric evaluation of research results and productivity; here metadata are important while access to full text is secondary, marginal.

On the other hand, some leaders of the open access movement and also governments and institutions began to distinguish between immediate and mandatory deposit of metadata and access to the full text, in order to accelerate the transition to green open access. Embargo periods are considered to be better than nothing and acceptable also because "immediate-deposit (...) and the contingency on eligibility for research assessment and funding (...)

ensure that the primary locus of deposit will be the institutional repository”⁸. Some initiatives such as the Open Access Button⁹ may help in coping with embargoed items.

For the scientist searching for documents on the Internet this lack of openness is less satisfying. Metadata without a link to the full text produce the same kind of barrier or pay wall as a publisher’s platform for authorized users only. The promise of access in some months or years (after the end of the embargo period) is not of interest to most scientific communities in need of recent publications, especially for those working in emerging and “hot”, cutting edge fields of research. This situation is even less satisfying because the largest part of lack of openness appears not to be due to embargo periods limited in time but to dissemination restricted to on-campus or institution-wide availability, without any specified time limits.

So, is the glass half empty or half full? Is this lack of openness a transitory effect, a kind of collateral damage of institutional decisions, individual choices, political strategies and intellectual property laws that will disappear with the advent of full open access? Or on the contrary is it (and will it remain) a basic feature of the new and complex cohabitation of institutional repositories and commercial publishing? The future will tell. In the meantime, what should be expected is that institutions will clarify and be explicit about their open access policy and assure the same level of quality for repositories as they have always done for their catalogues and databases.

Bibliography

I. Bernal (2013). ‘Open Access and the Changing Landscape of Research Impact Indicators: New Roles for Repositories’. *Publications* 1(2):56-77. [doi:10.3390/publications1020056](https://doi.org/10.3390/publications1020056)

L. Carr, et al. (2008). ‘Institutional Repository Checklist for Serving Institutional Management’. In *Third International Conference on Open Repositories 2008, 1-4 April 2008, Southampton, United Kingdom*. <http://pubs.or08.ecs.soton.ac.uk/138/>

F. Friend (2011). ‘Open Access Business Models for Research Funders and Universities’. Report, Knowledge Exchange, Copenhagen. http://www.knowledge-exchange.info/Files/Filer/downloads/Open%20Access/KE%20Briefing%20paper_BUSINESS_MODELS_FOR_RESEARCH_FUNDERS_AND_UNIVERSITIES-08-09-11.pdf

P. Jain (2011). ‘New trends and future applications/directions of institutional repositories in academic institutions’. *Library Review* 60(2):125-141. <http://dx.doi.org/10.1108/00242531111113078>

C. A. Lynch (2003). ‘Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age’. Report 226, ARL Association of Research Libraries. <http://www.arl.org/storage/documents/publications/arl-br-226.pdf>

J. Schöpfel & D. J. Farace (2010). ‘Grey Literature’. In M. J. Bates & M. N. Maack (eds.), *Encyclopedia of Library and Information Sciences, Third Edition*, pp. 2029-2039. CRC Press, London.

⁸ Comment of Stevan Harnad at <http://openaccess.eprints.org/index.php?/archives/1103-HEFCEREF-Adopts-Optimal-Complement-to-RCUK-OA-Mandate.html>

⁹ <https://www.openaccessbutton.org>

J. Schöpfel, et al. (2011). 'Open is not enough. A case study on grey literature in an OAI environment'. In *Thirteenth International Conference on Grey Literature: The Grey Circuit. From Social Networking to Wealth Creation. Washington, DC, 5-6 December 2011*, pp. 75-86, Amsterdam. TextRelease. http://archivesic.ccsd.cnrs.fr/sic_00908862

J. Schöpfel & H. Prost (2013a). 'Degrees of Secrecy in an Open Environment. The Case of Electronic Theses and Dissertations'. *ESSACHESS - Journal for Communication Studies* 6(2 (12)). <http://www.essachess.com/index.php/jcs/article/view/214>

J. Schöpfel & H. Prost (2013b). 'Back to Grey. Disclosure and Concealment of Electronic Theses and Dissertations'. In *GL15 Fifteenth International Conference on Grey Literature. The Grey Audit: A Field Assessment in Grey Literature. CVTI SR, Bratislava, Slovak Republic, 2-3 December 2013*. http://archivesic.ccsd.cnrs.fr/sic_00944662/fr/

K. Smith (2008). 'Institutional Repositories and E-Journal Archiving: What Are We Learning?'. *Journal of Electronic Publishing* 11(1). <http://dx.doi.org/10.3998/3336451.0011.107>

P. Suber (2012). *Open access*. MIT Press, Cambridge, MA. <http://mitpress.mit.edu/books/open-access>

A. Swan & C. Awre (2006). 'Linking UK Repositories: Technical & Organisational Models to Support User-Oriented Services Across Institutional & Other Digital Repositories'. Report, JISC, London. <http://www.jisc.ac.uk/publications/reports/2006/linkingukrepositoriesreport.aspx>

J. Willinsky (2005). *The Access Principle: The Case for Open Access to Research and Scholarship*. MIT Press, Cambridge MA.

Appendix – List of surveyed repositories

Europe

Chalmers <http://publications.lib.chalmers.se/> Chalmers Publication Library contains research publications produced at Chalmers University of Technology, Göteborg, Sweden.

CNRS <http://hal.archives-ouvertes.fr> HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research papers, including nearly 100 institutional repositories from French HE and research institutions.

CSIC <http://digital.csic.es/> Digital. CSIC the institutional repository of the Spanish National Research Council (CSIC)

Frankfurt a. M. <http://publikationen.ub.uni-frankfurt.de> publication server of Goethe University Frankfurt am Main

Geneva <http://archive-ouverte.unige.ch/> Open Archive UNIGE (university of Geneva)

Ghent <https://biblio.ugent.be/> Ghent University Academic Bibliography

INRA <http://prodinra.inra.fr> ProdINRA institutional repository of the French national agricultural research institute

KNAW <http://depot.knaw.nl> *Repository of the Royal Netherlands Academic of Arts and Sciences*

Milan <http://air.unimi.it> AIR *Archivio Istituzionale della Ricerca* of the University of Milan

Torino <http://porto.polito.it> PORTO open repository of publications produced by the scientific community of Politecnico di Torino

Uppsala <http://uu.diva-portal.org> institutional repository of the Uppsala University

Australasia

Macquarie University Sidney <http://www.researchonline.mq.edu.au> *ResearchOnline* open access digital collection

Monash University Melbourne <http://arrow.monash.edu.au> *Arrow* research repository

Queensland University of Technology Brisbane <http://eprints.qut.edu.au/> *QUT ePrints Archive*

RMIT Royal Melbourne Institute of Technology <http://researchbank.rmit.edu.au/> *RMIT Research Repository*

Swinburne University Melbourne <http://researchbank.swinburne.edu.au> *Swinburne Research Bank*

Asia

Dokuz Eylül University Izmir <http://deu.mitosweb.com/> open archive

Chiba University <http://mitizane.ll.chiba-u.jp/curator/> *CURATOR* Chiba University's Repository for Access To Outcomes from Research

University of Hong Kong <http://hub.hku.hk/> *HKU Scholars Hub* institutional repository

Singapore Management University <http://ink.library.smu.edu.sg/> *InK Institutional Knowledge* at Singapore Management University

National Taiwan University <http://ntur.lib.ntu.edu.tw/> *NTUR National Taiwan University Repository*

America (North, Central and South America, Caribbean)

University of Massachusetts Amherst <http://scholarworks.umass.edu/> *ScholarWorks@UMass* institutional repository Amherst

Western Kentucky University Bowling Green <http://digitalcommons.wku.edu/> *TopScholar* institutional repository

University of North Texas Denton <http://digital.library.unt.edu/> *UNT Digital Library*

All websites were accessed in March and April 2014.