Directories of Institutional Repositories: Research Results & Recommendations

Kathleen B. Oliver, MSLS, MPH
Johns Hopkins University

Robert Swain, BSN, MLS
Johns Hopkins University

Meeting:	**151 Information Technology with National Libraries with Academic and Research Libraries and Knowledge Management (part 1)**
Simultaneous Interpretation:	Yes

Abstract

At its 2005 business meeting in Oslo, the Health and Biosciences Libraries Section (HBLS) agreed that an international directory of institutional repositories would be a useful tool for IFLA. Members suggested that it could be mined and monitored for growth in numbers of repositories, their collections and content development, the services they provide, their acceptance and use by scholars, and their impact on scholarship. With that in mind, HBLS funded Johns Hopkins to 1) identify existing directories, and, for those found, 2) to describe their scope, record structure and updating mechanisms. In this paper, we will describe the results of our research. One directory, the University of Nottingham's OpenDOAR, stands out as the leader among the directories identified, particularly for the purposes envisioned at the Section's 2005 business meeting. This paper will describe and compare the scope, structure and update methodology of OpenDOAR and 23 other directories of institutional repositories, with particular attention to the health sciences. Based on our findings, we will offer suggestions for how the HBLS and IFLA might support an international directory of institutional repositories and how such a directory might be used for the advancement of scholarship globally.

Introduction

Institutional repositories are a part of an emerging movement towards open access to research information. At the 2005 IFLA meeting, the Health and Bioscience Committee co-sponsored a full-day Open Access session. Most interesting in this informative session was the discussion of the scope of scholarly publishing. David Prossner of SPARC Europe emphasized in his presentation the strategic link between institutional repositories and open access journals. Mr. Prossner described the four functions of scholarly publishing as: F1) Registration, F2) Certification, F3) Awareness and F4) Archiving. He stated that in the print environment, journals integrate these functions. In an electronic environment, the Budapest Open Access Initiative proposes two complementary strategies: self archiving and open access journals. He defined the essential attributes of institutional repositories and open archives as R1) institutionally defined, R2) scholarly in content, R3) cumulative & perpetual in nature, and R4) interoperable and openly accessible. Institutional repositories address publishing functions F1, F3 and F4. Open Access Journals address function F2, certification.

That said, the challenge remains in a number of areas related to the emergence of this new publishing model. One challenge is to monitor and support its progress and to identify and address important issues related to its development. This paper describes one approach to doing so: the development and support of a directory of institutional repositories, one international in scope reflecting the international nature of research and scholarship.

At its 2006 business meeting, the IFLA Health and Biosciences Section agreed that an international directory of institutional repositories would be a useful tool both for the Section and for IFLA as a whole. Members suggested that it could be mined and monitored for growth in numbers of repositories, their collections and content development, the services they provide, their acceptance and use by scholars, and their impact on scholarship. With that in mind we set out to 1) identify existing directories, and, for those found, 2) to describe their scope, record structure and updating mechanisms. In this paper, we describe the results of our research. One directory, the University of Nottingham's OpenDOAR, stands out as the leader among the directories identified, particularly for the purposes envisioned at the Section's 2005 business meeting. This paper will describe and compare the scope, structure and update methodology of OpenDOAR and 23 other directories of institutional repositories, with particular attention to the health sciences. Based on our findings, we will offer suggestions for how the HBLS and IFLA might support an international directory of institutional repositories and how such a directory might be used not only for the advancement of scholarship globally, but as a resource to measure that advancement.

Methods

We used the following methodology to identify directories of repositories listed in Table 2. All searches were conducted using the Google search engine. Searches strategies included "OAI", "Open Access Repository", "Repository Directory", as well as others. When viable sites were found they often linked to other relevant resources. These links were followed in an attempt to identify as many directories as possible.

2

Results & Discussion

Our research identified 23 directories of repositories. In Table 2, for comparison purposes, we describe for each directory identified: the title, internet address (URL), descriptive fields (metadata) for each record, the topical scope, its updating mechanism, the number of repositories, and notes on interesting features. Bio- and health science coverage was also noted. A few directories had features worth noting either from a functional perspective or because of their relationship to health and biosciences:

- ROAR, a registry of open access repositories, includes descriptive records for 610 repositories. Each descriptive record includes the country, the software used, and an OAI record count with a graph and screenshot. Its updating mechanism is a registration form but is also pings known sites for new additions. ROAR allows browsing by country, software type and content type.
- OAIster lists 611 repositories. It has a search function that permits the user to search all repositories at once. We couldn't find a way to limit search to just the bioscience or health field.
- Experimental OAI Registry at UIUC is a directory of 1047[1] engineering repositories. It updates its listings via email, and pings repositories to check if they are responding. This registry allows the user to generate reports based on multiple criteria.
- The Directory of Mathematics Preprint and e-Print Servers offers descriptive fields for address, contact information, special features, and remarks for each included repository. This directory lists umbrella servers that aggregate information from multiple repositories.
- The e-Print Network, produced by the U.S. Department of Energy (DOE) includes repositories with a broad coverage of subjects, and includes a category for biology and medicine. It lists individual scientists and is full text searchable.
- A number of the directories are software specific. These include: Digital Commons, Eprints, DSpace, and Fedora.

One directory, the University of Nottingham's OpenDOAR, stands out as the leader among the directories identified, particularly for the purposes envisioned at the Section's 2005 business meeting. It is international in scope. Subject coverage is noted and it is possible to browse and retrieve repositories with health and bioscience content. The analysis of this paper will describe a couple of approaches to mining OpenDOAR for information relevant to health and bioscience, illustrating how such a directory, to the extent that is comprehensive, might be used as a tool to measure development of this emerging model for scholarly exchange, and, possibly, the distribution of innovation globally.

A directory such as OpenDOAR makes it easier to identify and mine the individual repositories. An example of this can be found in Table 1 which describes the number of repositories in total and by country that have some content in the health and biosciences. These data were retrieved using the browse feature available in OpenDOAR. They could be graphed over time as a visual measure of growth of participation in the directory. If the directory is well supported and matures as a reasonably comprehensive resource, such a graph might be seen more as a measure of growth in numbers of repositories, not just participation in the directory.

Because it is possible in OpenDOAR to monitor repository growth geographically, another area of research, made more accessible by OpenDOAR or a directory like it, is the

---

[1] As of early March 2006

relationship of repository content in health and biosciences to the research and development investment by national governments. From there it may be possible to monitor the growth and distribution of innovation geographically around the world. Science and Engineering Indicators 2006 published by the U.S. National Science Foundation states:

Increasingly, the international competitiveness of a modern economy is defined by its ability to generate, absorb, and commercialize knowledge. Most nations have accepted that economic policy should focus not only on improving quality and efficiency but also on promoting innovation. Absolute levels of R&D expenditures are important indicators of a nation's innovative capacity and are a harbinger of future growth and productivity. Indeed, investments in the R&D enterprise strengthen the technological base on which economic prosperity increasingly depends worldwide. The relative strength of a particular country's current and future economy and the specific scientific and technological areas in which a country excels are further revealed through comparison with other major R&D-performing countries.

It makes sense that scholarly output correlates to R&D investment (Figure 1 and 2). It would prove interesting to compare health and bioscience repository content (together with other science categories) to R&D expenditures (as a percentage of Gross Domestic Product (GDP)) to see if a relationship could be defined and what the nature of it might be. Certainly scholarly output stored in repositories, together with published literature, serves as one measure of innovation. A directory such as OpenDOAR facilitates the tracking of that geographically.

A cursory review of R&D investment in the context of currently registered repositories (Table 1) does not suggest an obvious correlation of the two numbers. With a well developed and reasonably comprehensive directory, it is reasonable think a relationship would emerge. If a relationship between scholarly content stored in repositories and R&D expenditures could be established or defined, over time one could look at it to track shifts in the geographic distribution of innovation. Or, one could hypothesize that, if growth rates in repositories or repository object volume reflect those in R&D over time, such parallelism suggests repositories have reached a level of acceptance within the scholarly community. Any way one looks at the data, it would be interesting to track these figures over time and investigate their relationship. Again, directories like OpenDOAR, if well supported, would make such research and analysis more possible.

Summary

In 2005, the IFLA Health and Bioscience Section concluded that a directory of institutional repositories would be a useful tool in monitoring the success of one aspect of the new publishing model proposed by Prossner. The Section set out to identify directories of repositories, and, if possible, find one that was international in scope and included those with health and bioscience content. A search identified twenty four directories. We prepared a summary table where we noted scope, record structure and updating mechanisms for each directory found (Table 2). Based on the review of these, OpenDOAR emerged as the clear leader in terms of scope and usability for the purposes envisioned by the Health and Biosciences Section. We explored the possibility of using OpenDOAR to monitor trends and global shifts of centers of innovation. We suggested if a directory of repositories were reasonably comprehensive and included suitable meta-data, that it would facilitate further exploration of the relationship between scholarly output stored in repositories and R&D expenditures, an accepted indicator of innovation. These findings suggest that investment and support of OpenDOAR, or a similar directory, could be productive not only in facilitating

4

access to scholarly output and the advance of repository functionality but also in monitoring its acceptance by scholars and as a tool to measure innovation.

The utility and success of OpenDOAR would require that it be comprehensive and financially viable. With recognition of the importance of OpenDOAR as a resource, IFLA member countries, libraries and institutions are well positioned to advocate and ensure registration of their repositories in OpenDOAR. IFLA could advocate and assist in developing mechanisms to ensure the financial viability of OpenDOAR. If OpenDOAR achieved financial viability and approached comprehensiveness as an international database of repositories, it could be exploited in a number of ways. A few examples might be:

- International and national agencies could fund research using OpenDOAR as a resource; such research could identify trends in scholarship and innovation globally, in a useful and unique manner. Other hypotheses related to scholarship could be generated and tested using this resource.

- OpenDOAR could be used as a resource to develop themes for IFLA and other professional and scholarly forums whose purpose is to advocate, develop and monitor open access to scholarship and its consequences.

- It could facilitate the functional development of both repositories and of OpenDOAR, as a directory of repositories, by making it easier to find and share innovation in updating technology, search and retrieval technology, and data collection and analysis functions. For example, it would be useful if the directory had the capacity to conduct sophisticated searches across repository content, a pro-active update mechanism to ensure the currency and completeness of the directory, and functions that would enable users to quantify scholarly objects across repositories.

These are but a few examples of how the HBLS and IFLA might support an international directory of institutional repositories and how such a directory might be used not only for the advancement of scholarship globally, but as a resource to measure that advancement.

References

U. S. National Science Foundation. Science and Engineering Indicators 2006. Chapter 4: Research and Development: Funds and Technology Linkages.
http://www.nsf.gov/statistics/seind06/c4/c4s6.htm

Prossner, David. Fulfilling the promise of scholarly publishing: can open access deliver? In Proceedings of the IFLA 2005 Satellite meeting No 17. Open access: the option for the future!? Rikshospitalet University Hospital, Oslo, Norway: August 2005.
http://www.ub.uio.no/ifla/IFLA_open_access/programme_abstracts.htm

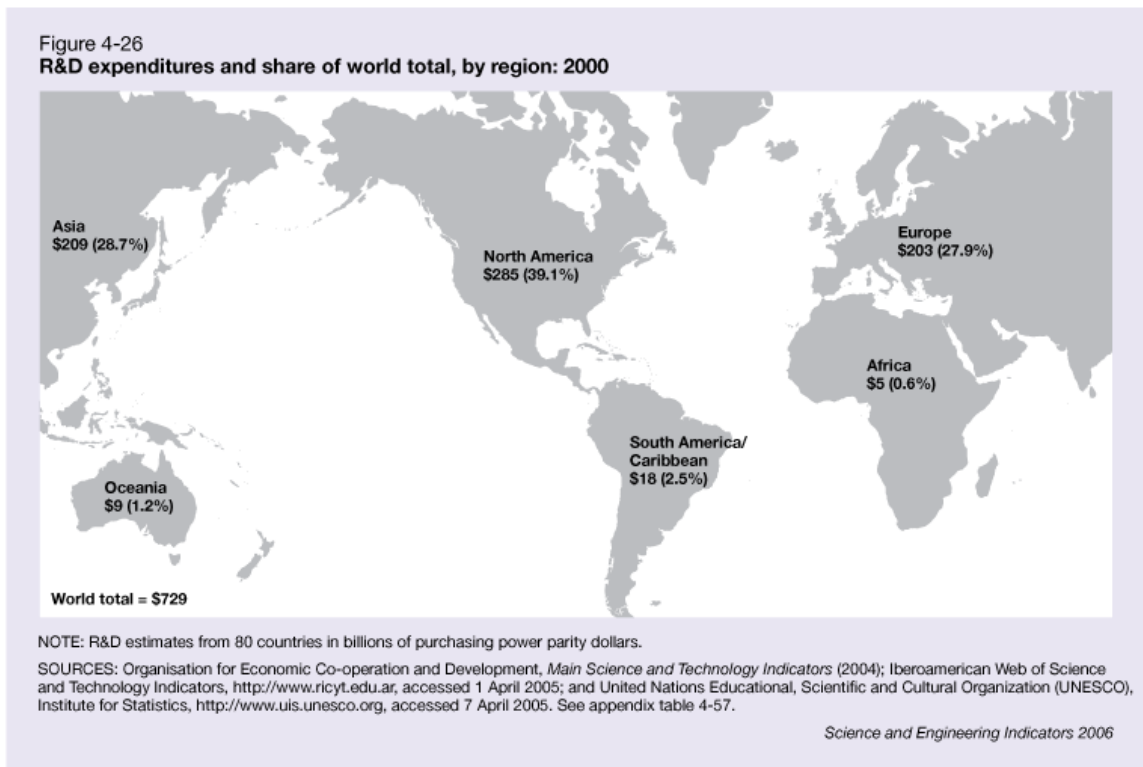Table 1. Repositories & Repositories With Health/Science Coverage[2]

	Repositories	With Bioscience Content	With Health Content	R&D as a % of GDP[3]
All Countries	380	155	144	
Australia	17	9	12	1.54
Austria	3	1	1	2.19
Belgium	9	4	2	2.33
Brazil	10	5	5	1.04
Canada	18	5	8	1.87
Chile	2	0	2	0.57
China	2	1	1	1.22
Columbia	2	1	0	0.10
Denmark	3	3	0	2.52
Finland	3	2	2	3.46
France	21	10	5	2.26
Germany	39	18	16	2.50
Greece	1	0	0	0.65
Hungary	3	0	0	0.95
India	8	3	1	N.D.[4]
Ireland	2	1	0	1.13
Israel	1	0	0	4.90
Italy	15	4	3	1.11
Japan	4	2	2	3.12
Mexico	1	0	0	0.39
Namibia	1	0	0	N.D.
Netherlands	14	5	8	1.88
New Zealand	1	1	1	1.16
Norway	3	2	2	1.67
Pakistan	1	0	0	N.D.
Portugal	3	1	0	0.94
Russia	1	0	0	1.28
Singapore	1	0	0	2.15
Slovenia	1	0	0	1.53
South Africa	4	3	3	N.D.
South Korea	0	0	0	2.64
Spain	5	3	3	1.03
Sweden	18	9	11	4.27
Switzerland	4	2	2	2.57
United Kingdom	56	23	18	1.87
United States	102	35	34	2.67
Venezuela	2	1	2	N.D.

Figure 1. From Science and Engineering Indicators 2006 published by the U.S. National Science Foundation

Figure 4-26
R&D expenditures and share of world total, by region: 2000

Asia
$209 (28.7%)

North America
$285 (39.1%)

Europe
$203 (27.9%)

Africa
$5 (0.6%)

South America/
Caribbean
$18 (2.5%)

Oceania
$9 (1.2%)

World total = $729

NOTE: R&D estimates from 80 countries in billions of purchasing power parity dollars.

SOURCES: Organisation for Economic Co-operation and Development, *Main Science and Technology Indicators* (2004); Iberoamerican Web of Science and Technology Indicators, http://www.ricyt.edu.ar, accessed 1 April 2005; and United Nations Educational, Scientific and Cultural Organization (UNESCO), Institute for Statistics, http://www.uis.unesco.org, accessed 7 April 2005. See appendix table 4-57.

Science and Engineering Indicators 2006

Source Links: Close window
Source data

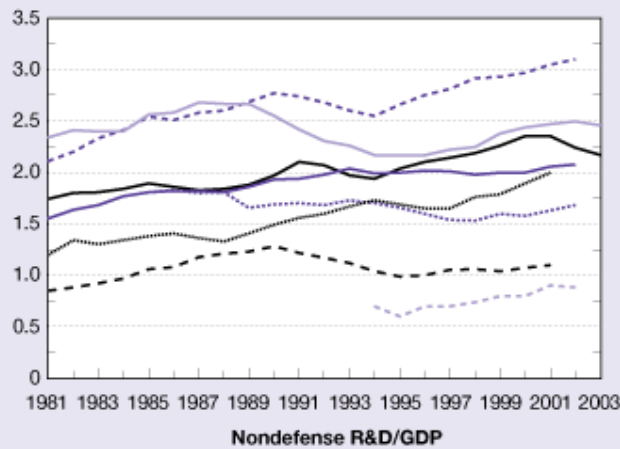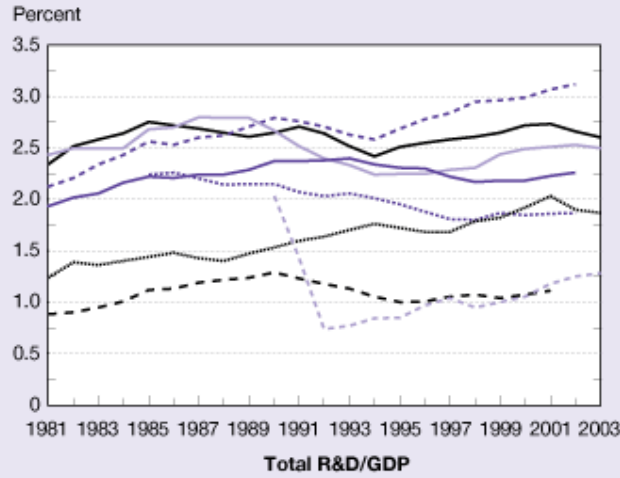**Figure 2. From Science and Engineering Indicators 2006 published by the U.S. National Science Foundation**

Figure 4-30
R&D share of gross domestic product, by selected countries: 1981–2003

United States · · · · · · United Kingdom
- - - - Japan - - - - Italy
Germany · · · · · · Canada
France · · · · · · Russia

Percent

Total R&D/GDP

Nondefense R&D/GDP

GDP = gross domestic product

SOURCE: Organisation for Economic Co-operation and Development, *Main Science and Technology Indicators* (2004). See appendix tables 4-42 and 4-43.

Science and Engineering Indicators 2006

Table 2. Online International Directory of Institutional Repositories[6]

Title	Link	Metadata	Scope	Update Mechanism	# listed	Notes
The Directory of Open Access Repositories - OpenDOAR	http://www.opendoar.org/	Country Organization Subjects Type OAI base URL Description	Very broad scope.	Repository registration, not an OAI registration site.	347	Allows browsing by country, content type, or subject. Health Sciences is listed and contains 134 repositories. Funding extends to mid 2006.
Open Archives Initiative - Repository Explorer	http://re.cs.uct.ac.za/	OAI Testing Site	Used for testing OAI compliance	None listed	>100	This site is designed for testing OAI sites.
SPARC	http://www.arl.org/sparc/repos/ir.html	Content System software Contact information Link	Institutional repositories, excludes discipline specific servers.	Email address for additions and corrections	42	Repositories are listed by country of origin.
Registry of Open Access Repositories (ROAR)	http://archives.eprints.org/	Country Software OAI record count with graph Description	Very broad scope	Registration form. Also pings known sites for new additions	610	Allows browsing by country, software type, content type. Includes a graph showing # of OAI records plus a screen shot.
OAIster	http://oaister.umdl.umich.edu/o/oaister/viewcolls.html	Link Description OAI record count	Very broad	Registration information	611	Also has a search feature that allows the user to search all repositories at once. No way to limit search to just the Health Science field.
Celestial	http://celestial.eprints.org/cgi-bin/status	Link Namespace Harvest Method Records Error date	Very broad	None Listed. Pings repositories to check for additions and errors.	928	A metadata harvest site

---

[6] As of March 1, 2006

Experimental OAI Registry at UIUC	http://gita.grainger.uiuc.edu/registry/	Link Protocol version Date last checked	Engineering	Email. Pings repositories to check if they are responding	1047	Allows user to generate reports based on multiple criteria.
Open Archives	http://www.openarchives.org/Register/BrowseSites	Link OAI information Repository identifier	Broad	Registration page	401	Contains an XML formatted list of all repositories.
OpCite	http://opcit.eprints.org/explorearchives.shtml	Link Description				Last updated on June 30, 2003
Directory of Mathematics Preprint and e-Print Servers	http://www.ams.org/global-preprints/index.html	Link Contact Special Features Remarks	Mathematics	None listed	-	Lists umbrella servers that aggregate information from multiple repositories
E-print Network, U.S. Department of Energy (DOE)	http://www.osti.gov/eprints/	Lists individual scientists	Very broad, includes a category for Biology and Medicine	Distributed Explorit and Explorit Focused Crawler by Deep Web	Very large	Run by Department of Energy. Links to individual researcher works in repository. Full text searchable
Virtual Technical Reports Center: EPrints, Preprints, & Technical Reports on the Web	http://www.lib.umd.edu/ENGIN/TechReports/Virtual-TechReports.html	Link	Broad -	Email		Long list, some dead links or links where authentication is required
Open Access Webliography	http://www.escholarlypub.com/cwb/oaw.htm#e-prints	Directory of repository directories	Broad	None		Only a small subset related to repositories. Open Access Webliography
DigitalCommons@The Texas Medical Center	http://digitalcommons.library.tmc.edu/about.html	Links	Digital Commons sites	Email	53	Lists a Johns Hopkins repository http://digitalcommons.dkc.jhu.edu/

Eprints Archive	http://www.eprints.org/software/archives/	Link Number of resources	Eprints sites	Derives information from http://archives.eprints.org/	196	Harvests list from http://archives.eprints.org/
The Association of Learned and Professional Society Publishers	http://www.alpsp.org/htp_openarc.htm	Link	Varied		10	Short list
Digital Commons	http://www.umi.com/products_umi/digitalcommons/	Link	Digital Commons sites	None given	44	Digital Commons site
D Space	http://wiki.dspace.org/DspaceInstances	Link	DSpace sites	Wiki interface for adding repositories	130	DSpace sites
Fedora	http://www.fedora.info/community/	Link	Fedora sites	None listed	20	Informational site about Fedora
SDL : Search Digital Libraries	http://drtc.isibang.ac.in/sdl/archives.php	Link Record number	Varied	Not listed. Pings repositories for number of records	13	Search interface and data harvester
OAI Scotland Information Service	http://hairst.cdlr.strath.ac.uk/oaisis/Repositories.htm	Link Institution	Repositories in Scotland	Not listed	8	Provides information about setting up a repository
Public Knowledge Project, Open Archives Harvester	http://pkp.sfu.ca/harvester/archives.php	Link Record number	Very broad	Not Listed	250	Search interface and data harvester

Digital Academic Repositories	http://www.darenet.nl/en/page/language.view/repositories	Link	Universities in the Netherlands	Email	17	Lists repositories as well as individual researchers
Open Language Archives Community	http://www.language-archives.org	Very detailed http://www.language-archives.org/archive.php4?id=3	Language oriented repositories	Form	34	Directory contains a report card for each repository. http://www.language-archives.org/tools/reports/archiveReportCard.php?archive=3