

## **Institutional Repositories and Open Access**

Thaddeus Lipinski  
Web Applications Manager  
Bodleian Libraries  
University of Oxford

Published: November 2011  
Multimedia Information and Technology , Vol 37, No 4, November 2011  
<http://mmitjournal.org.uk/>

### **Introduction**

Within the last two decades, a new form of scholarly publishing has developed within educational organisations. In addition to the usual process of submitting papers for print publication, academics and researchers can also self-archive their work by depositing it in an Institutional Repository (IR). Institutional Repositories are open access archives for papers, articles, theses or other material which are hosted and maintained by the institution with associated policies of digital archiving and preservation.

An early description of IR's "capturing and preserving the intellectual output of a single or multi-university community" [Crow] still holds. IR's provide content that can supplement print publications, and allows academics globally to share research material or other resources that may otherwise have remained dormant in local computer files or had limited distribution.

At the moment effort and staff time is needed to advocate, collaborate, define policies, populate and maintain an open access repository, so the user can find information simply. There is a trend to merge existing small-scale archives into a single institution wide-one. For example, the University of Oxford's Bodleian Libraries maintains the Oxford University Research Archive (ORA), which is also the mandated deposit location for e-theses. Other departmental repositories within Oxford University such as Mathematics are now moving over their contents to ORA. ORA is becoming the principal institutional gateway for Oxford University.

There is another aspect to IRs. Typically, universities already have an archivist for its university archives. This differs from IR's in that university archivists determine what is archived (which can also be physical material) whereas institutional repositories aim to capture and preserve the research output. IR's can also store other content for dissemination, such as support staff lectures slides and other grey literature.

### **Some History**

The first recognised online repository is arXiv (pronounced archive) developed for science e-prints at the Los Alamos National Laboratory, USA in 1991. arXiv is now maintained by Cornell University. In the UK CogPrints was set up in 1997. CogPrints is hosted by the University of Southampton and focuses mainly on papers in Psychology, Neuroscience and other fields.

Considerations about enabling services to operate together in a seamless manner led to The Open Archives Initiative in 1999. Then in 2003 the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities was signed, “to promote the Internet as a functional instrument for a global scientific knowledge base.”

Eprints software was developed in 2000 for institutional repositories, and the SHERPA project in the UK helped establish HE e-print servers. Open source architecture designed specifically for repositories such as DSpace and Fedora lowered the barriers to implementation.

In the UK, Southampton University organised a conference about the Berlin Declaration in 2005. A workshop discussing repositories had been held a week earlier. These events seem to have spurred momentum and more IR's appear from around 2005/2006.

The European Commission funded SOAP project (Study of Open Access Publishing), presented its results on January, 2011 in Berlin. The project which surveyed researchers in Europe analyses the open access publishing landscape of 2010, for both publishers and researchers.

[Image I: Growth of the OpenDOAR Database - United Kingdom

Ref: <http://www.opendoar.org/> Accessed 28 May 2011]

## **Benefits**

For an academic or organisation, IR's can provide a place to deposit open access research articles or coursework material, as well as improve the availability of quality research publications worldwide, with benefits to developing countries. For an academic or researcher it can help a market their work and using the retrieval statistics provided by IR's, it also helps assess the impact of the research. Conversely, students looking to publish can use the quality of articles deposited in IR's and their associated journals to decide in where to formally publish. Not all research deposited into IR's need be in English, of course.

## **Issues to Consider**

### Policies

The IR manager may need to consider policies such as appropriate metadata records; policies dealing with mandated or voluntary submission; copyright restrictions and licensing arrangements. Other concerns relate to legal retention and replacement of archived documents; what level to moderate; a collections policy; training or statistical information to show.

### How to deposit

How research material is imported into a repository can be left to the repository team. Authors can self-archive their material, or it can be mediated, either by an administrative staff associated with an

academic department or - more likely - by staff based with the IR service. The records actually held in a repository may be only metadata associated with external material, or the full text.

### Licensing Issues

In many cases, the repository may require the copyright holder to grant the institution a non-exclusive licence. This licence is intended to give the institution, and by extension the repository, the right to store, copy or modify the material as part of the preservation and dissemination process. The licence may be mandated for academics or allow an opt-out option. Alternatively, material can be published in the repository under a Creative Commons license.

Material may be deposited as pre-prints – work in progress – or published articles: post-prints. A common stated concern for academic is if a journal does not allow free public distribution, it will limit the author's publication options. One way to avoid this dilemma may be to embargo access to an article in the IR for a period, to allow a journal to publish it first. This could also occur with dissertations.

### **Open Access**

Underlying the concepts of IRs are concepts such as 'open access' and 'archive'. Open access refers to the lack of restrictions imposed on accessing the information, such as subscriptions. In this sense the data is open – open for access. The Open Archives Initiative (OAI) is a major forum that develops and promotes the interoperability standards to enable access to content held in repositories.

What does interoperability mean? It refers to the mechanisms underlying the various repositories such as metadata formats, their underlying architecture and integration with established scholarly communication conventions.

Within the Open Archives community, the term 'archive' is generally accepted as a synonym for repository of scholarly papers, rather than the connotations of long term storage, preservation and authorisation.

### **How does it Work?**

An IR can be viewed as a data provider. When material is deposited in an IR, metadata in XML format is created for it. This metadata conforms to published standards, for example Dublin Core. The metadata can then be *harvested* by service providers. A distinct service provider can use the metadata to build added value services. OIA has developed a protocol for harvesting data, known as OAI Protocol for Metadata Harvesting (OAI-PMH). This protocol allows for the sharing and syndication of metadata between services.

[Image II: How services can harvest from multiple Data Providers]

The principle of data harvesting is not new; existing search tools allow subscription service providers to unify disparate online collections into a single point of entry, and ubiquitous Google allows simplistic keyword searching over web documents. What *is* new is the principle applied to formerly unavailable or restricted scholarly output with the emphasis on open access. In addition, a simple interface allows scholars to deposit material and researchers to discover it.

In the past, repositories were difficult to install and use – researchers needed to learn each repository’s quirks. Nowadays global repositories can be queried without the user needing to understand the underlying architecture. Software specifically designed for IR’s and interoperable protocols mean that attention can be turned to the other aspects of hosting a repository already mentioned.

### **Service Providers for IR’s**

There exist directories of global IR’s and service providers. Some are: listed below.

*OpenDOAR* is an international directory of academic open access publication repositories with full text resources. It is maintained by SHERPA Services, University of Nottingham, whose institutional members have a variety of repository environments.

Registry of Open Access Repositories (ROAR) is an international directory developed by the EPrints team at University of Southampton. The register usefully indicates the repositories’ growth and size.

Repository 66 is mashup based on ROAR and *OpenDOAR*, which maps the worldwide locations of open access repositories.

OpenDepot enables academics worldwide to enter their research output into a suitable IR. It lists those universities and organisations which have an online repository. For those without a local repository, including unaffiliated researchers, OpenDepot is a place of deposit, available for other service providers to harvest.

A UK Institutional Repository Search is maintained by MIMAS.

OAIster is a union catalogue harvested from worldwide open access resources.

BASE, operated by Bielefeld University Library, is an example of a service provider that presents its interface in several languages.

Google Scholar presents the content of open access repositories as well as standard websites.

### **Conclusion**

Many countries are involved in the development of repositories, and the number is growing. Looking at ROAR and Repository 66 in 2011 we see that Western Europe and North America have the largest number of repositories, followed by Japan and Australia. Other countries also have a

major presence e.g. Brazil and India. However a number by itself is no indicator of the quality of resources “out there” and many open access repositories may be small or subject specialist ones.

Libraries are well suited to developing the repository, with help from the IT sector. Institutional libraries already maintain online web portals to resources and catalogues and manage metadata on a daily level. With budget cuts taking place and journal subscription rates rising, the creation of institutional repositories which benefit academics - and other organisations - will increase.

Worldwide, it will also help libraries become the focus point for digital preservation.

## References

BASE <http://www.base-search.net/>

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

Berlin 3 Meeting <http://www.eprints.org/berlin3/program.html>

Chanson, H. *Some Experience with Digital Publishing, Open Access Repository, Research Impact and Quality* [http://espace.library.uq.edu.au/eserv/UQ:121513/Oar\\_For2.pdf](http://espace.library.uq.edu.au/eserv/UQ:121513/Oar_For2.pdf)

Crow, R. *The Case for Institutional Repositories: a SPARC Position Paper* [http://www.arl.org/sparc/bm~doc/ir\\_final\\_release\\_102.pdf](http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf)

EPrints <http://www.eprints.org/>

Google Scholar <http://scholar.google.co.uk/>

Harnad, S. *For Whom the Gate Tolls?* <http://users.ecs.soton.ac.uk/harnad/Tp/resolution.htm>

MIMAS <http://mimas.ac.uk/>

OIAster <http://www.oclc.org/oaister/>

Open Access Initiative <http://www.openarchives.org/>

OpenDepot.org <http://opendepot.org/information.html>

OpenDOAR <http://www.opendoar.org/>

Oxford University Research Archive (ORA) <http://ora.ouls.ox.ac.uk/>

Protocol for Metadata Harvesting (OAI-PMH) <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Registry of Open Access Repositories (ROAR) <http://roar.eprints.org/>

Repository 66 <http://maps.repository66.org/>

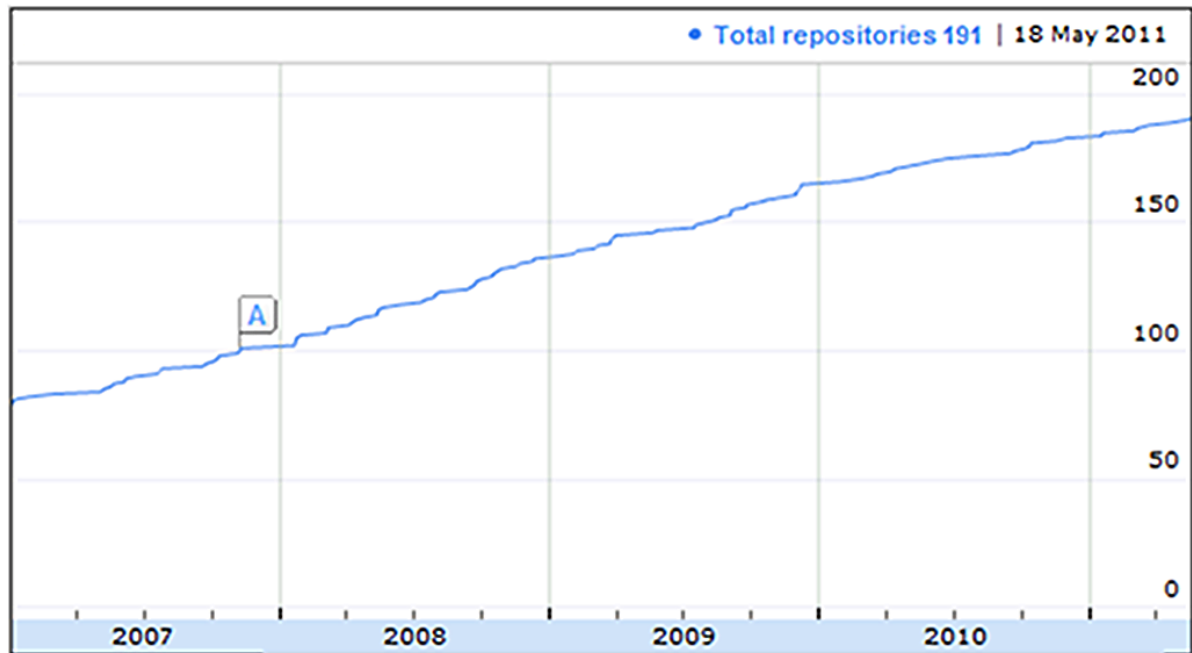
SHERPA <http://www.sherpa.ac.uk/>

SOAP Report <http://project-soap.eu/report-from-the-soap-symposium/>

UK Institutional Repository Search <http://irs.mimas.ac.uk/demonstrator/>

Woodward, H. and Estelle, L. (2010) *Digital Information: order or anarchy?*

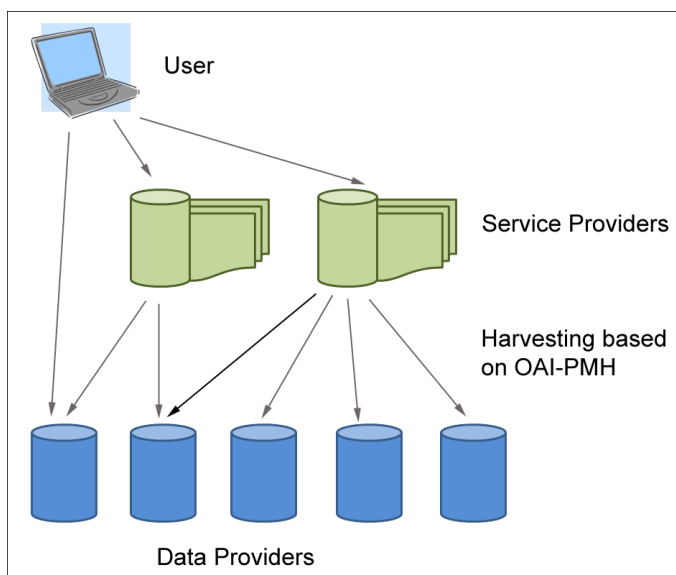
**Images:**



Growth of the OpenDOAR Database - United Kingdom

This graph shows the growth of UK repositories, as listed by the OpenDOAR Database. Label (A) indicates 100 repositories, reached on 9 November 2007.

Ref: Adapted from <http://www.openoar.org/> Accessed 28 May 2011



How services can harvest from multiple Data Providers

Ref: Adapted from <http://www.oaforum.org/tutorial/>