A Librarian's Guide to Institutional Repositories

Joanna Barwick (Pilkington Library, Loughborough University) and Miggie Pickton (University of Northampton Library)

Introduction

Institutional repositories (IRs) are a recent feature of the UK academic landscape. You may already have one at your workplace (in which case you might be better to skip to the next article); you will probably have heard the term being bandied about by your colleagues; you might even have come across one when trawling the Web. But what is an IR? Should your institution have one? And if so, how would you go about creating it? These are some of the questions we hope to address in this short article.

What is an Institutional Repository?

Foster and Gibbons (2005) describe an IR as

> "an electronic system that captures, preserves, and provides access to the digital work products of a community".

The IR is a digital archive, owned and maintained at either departmental or institutional level. Essentially, it is a tool for collecting, storing and disseminating information.

The content of an IR may be purely scholarly (Crow, 2002) or may comprise administrative, teaching and research materials, both published and unpublished. All types of digital product may be stored – articles, reports, presentations, images, data, even multi-media items. Importantly, the IR is cumulative and perpetual – it houses a *permanent* record of work.

Since a primary goal of an IR is to disseminate the institution's intellectual product, it is important that content is accessible both within and outside the host institution. In technical terms, it should be both open and 'interoperable'. In practice, this means that IR material should be described by metadata that can be harvested by external software. The Open Archive Initiative (OAI) exists to develop and promote the standards that will facilitate this. Its Protocol for Metadata Harvesting (OAI-PMH) enables the sharing of

metadata between services, and is the standard adopted by most IRs. Search engines such as OAIster (http://oaister.umdl.umich.edu/o/oaister/), ARC (http://arc.cs.odu.edu/), Citebase (http://www.citebase.org) and Google Scholar (http://scholar.google.com/) then find and enable the retrieval of IR material.

Why have an Institutional repository?

The impetus for IRs came from an increasing awareness that the products of publicly funded academic research are therefore 'public goods' (Berry 2000, p38) and as such, should be made freely available. The principle of 'Open Access' (OA) has received much attention in the literature, most recently following the publication of the UK House of Commons Science and Technology Committee's report "Scientific Publications: Free for all?"

(http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/399.pdf) and the subsequent Research Council UK's position statement (http://www.rcuk.ac.uk/access/statement.pdf). The RCUK propose that their award holders should be mandated to make their outputs available in OA format – either in OA journals or in a digital repository.

So the principle of OA has official blessing and the IR provides a means of supporting this. But does the IR offer any benefits to the more immediate stakeholders – the institution and the contributors? A survey of the OA literature suggests it does:

To the institution, an IR offers:

- A means of increasing visibility and prestige. A high-profile IR may be used to support marketing activities to attract high quality staff, students and funding.
- The centralisation and storage of all types of institutional output, including unpublished or 'grey' literature.
- Support for learning and teaching. Links may be made with the virtual learning environment and the library catalogue (Day 2003). Shared material may be 're-purposed' and reused.
- Standardisation of institutional records. The compilation of an 'institutional CV' (Swan et al. 2005b, p.8) and of individual online CVs linked to the full text of articles (Harnad et al 2003) are possible outcomes.

- Leverage of existing systems. By exploiting existing computer networks, IT services and library expertise, the IR enables these units to demonstrate greater efficiency (Yeates 2003, p.98).

- Improvements in administrative efficiency, especially if the IR is integrated with other institutional data management systems. Obligations regarding records management, health and safety record-keeping, and freedom of information may all be supported by the IR (Heery and Anderson 2005, p.5).

- Possible long-term cost savings. Some hope that the widespread adoption of IRs will ultimately enable savings to be made in subscriptions to academic journals. This however is unlikely to occur until a 'critical mass' of content is achieved (Pinfield 2002, p.262).

There are also benefits to authors:

- Increased dissemination and impact. Research has shown that the usage and citation of open-access material is greater than that of restricted access work (Antelman 2004, p.373, Kurtz 2004, p.1, and others).

- Storage and access to a wide range of materials, including digital representations of artwork, data sets, and audio-visual material. Compared with traditional print-based publication, the IR offers greater variety and flexibility; compared with personal or departmental websites, the IR offers greater security and longer term accessibility.

- Feedback and commentary. Some digital repositories permit the deposit of pre-publication 'preprints', enabling authors to assert priority and receive commentary.

- Provision of added value services such as hit counts on papers, personalised publication lists and citation analyses (Hubbard 2003, p.244, Pinfield 2002, p.262).

What are the snags?

Despite the clear benefits of IRs to both institutions and authors, the road to implementation has not always run smoothly. Some of the concerns raised have included:

- Cost. The existence of free open-source software for creating IRs has meant that initial financial costs may not be high (Steele 2003, p.3). Ongoing costs, however, especially staff costs (time spent drafting policies, arranging licensing agreements, developing guidelines, publicising the repository, training and supporting users and creating metadata), may be significant (Crow 2002, p.28, Horwood et al. 2004, p.174).

- Difficulties with generating content. A successful IR depends on the willingness of authors to deposit their work. Authors' existing working practices, and their attitudes and concerns, sometimes militate against this.

- Sustaining support and commitment. The IR is a long-term commitment. Its maintenance must be an institutional strategic goal. Methods of long term digital preservation are as yet untested.

- Rights management. Materials placed in an IR are subject to intellectual property rights. These may be owned by the institution, the author, or in the case of a postprint, a publisher (Gadd et al. 2003a, p.245). Despite clear evidence that many journals publishers support self-archiving (EPrints.org, 2005) concerns over intellectual property rights are a major deterrent to many authors (Heery and Anderson 2005, p.13, Pickton and McKnight 2006).

The dual challenges in implementing an IR are to promote the benefits it offers, while allaying stakeholders' concerns.

Case study: The Loughborough University Repository

At Loughborough, we took into account the issues outlined above when considering creating and maintaining an IR, and in 2004, we decided to go ahead. Our project began with the assembly of a committee to oversee the development of the IR. Clearly, the implementation of an IR requires a wide range of skills; skills that we, as information professionals, already had amongst our colleagues. By drawing upon the skills of these individuals, the IR Steering Committee has helped to ensure the healthy growth of the new service.

In June 2005, Jo Barwick began an appointment as Support Services Librarian at the Pilkington Library. In the first year of her post, she will be responsible for the day-to-day

coordination of development of the IR; with the view that, once established, the workload will be embedded into the general work of other Library staff.

Choosing the software

Under the guidance of our Systems Team Manager, Gary Brewerton, the different software options were investigated. There are now a wide range of open-source software products (the key players are E-Prints / DSpace/ Fedora); and there are some commercial options, for example BioMedCentral, as well as other packages being developed by library management systems companies. Open-source software is preferable (as it is free!); however, if your Library does not have the technical expertise in-house, a commercial package may be a better option. At Loughborough, we were fortunate to have sufficient technical support to opt for an open-source product, DSpace. This software offered a decent web interface yet still had the functionality to hold various file formats (including image and multi-media).

Gaining support (and funding)

It was crucial to our ongoing development to have support from a number of internal sources. Our University Librarian, Mary Morley and Support Services Manager, Jeff Brown, invested time in presenting the project to various university committees in the planning stages. This period was also used to identify 'early adopters' – departments that were happy to take part in the pilot stage of the service. (See Gathering Content, below)

Policy decisions

A number of things needed to be set in place before we started collecting material. We quickly established a structure for the collections within DSpace and made decisions on standards to ensure interoperability. (DSpace uses Dublin Core records and we have implemented LCSH.) We also drew up a licence for authors with the help of Steering Committee members, Lizzie Gadd and Charles Oppenheim. This licence was based upon the SHERPA model and Creative Commons.

Gathering content and advocacy

Having identified six supportive 'early adopters', from June 2005 we started working closely with these departments to source content. We targeted individuals who were already uploading their research to their personal web-pages and people publishing in IR-friendly journals. This resulted in an initial set of around 250 papers. The service has now been more widely publicised: with a view to launching the service formally in June 2006. We are working with our academic librarians and their departmental contacts to encourage others to take advantage of the service. In some cases, this has been very successful, but others have been slower to accept the principles of OA and the benefits of IRs.

Challenges of implementation

Convincing academics of the benefits of an IR has proved to be the project's major challenge. Many are highly sceptical and view this as another demand on their already limited time. At present, we are not asking academics to self-archive; instead we are doing this for them within the Library. It was hoped this approach would encourage them to participate more freely. Other academics are concerned about quality issues, or uncertain of our assurances that publishers will allow them to deposit their work. All of these issues involve patience and our highly-tuned negotiating skills!

One major problem we had not anticipated was which version of the material we were to use. Most publishers, although they will allow authors to archive their work on IRs, will not allow them to use the publisher-produced PDF. This means that we will often have to ask academics to supply us with their own final version, which has led to confusion: many academics do not keep their final version (they do not need to as the publisher sends them a pretty PDF); with others, their final version is so different to the actual published version, they are concerned about quality issues of archiving a pre-published version. Convincing them of the "Harnad/Oppenheim" view, that any copy is better than no copy, can be difficult. We are now encouraging authors to hold on to their final version in the hope that we can change behaviours. Time will tell…

Implementing an IR: recommendations

We recommend that anyone considering implementing an IR should take the following overlapping steps:

1. Conduct background research – including talking to the folk who have been through the process already
2. Establish agreement in principle from colleagues and departmental management
3. Gather a team of experts to draw upon (especially in the areas of technology, intellectual property, metadata, policy and advocacy)
4. Establish the principles which will underpin the IR
5. Recognise the resource implications (especially in staff time)
6. Win institutional support and commitment at the highest level
7. Identify short and long term sources of funding (sustainability is key)
8. Choose, acquire and install the software
9. Define IR policy and procedures (including content types and formats, task responsibilities, organisation of the IR, etc.)
10. Identify a group of sympathetic stakeholders with whom a pilot project may be undertaken
11. Conduct the pilot project
12. Review and refine IR policy and procedures
13. Know the answers – make sure your advocates are clear about the benefits of the IR and have solutions to all the potential objections
14. Proactively invite content from across the institution
15. Promote the IR relentlessly and tirelessly…

…then sit back and feel proud that you have contributed to the advancement of human knowledge.

Further information

To learn more about some of the concepts and issues raised in this article, please see the web sites below. Several of these also have links to other useful information.

EPrints (http://www.eprints.org/) and **DSpace** (http://www.dspace.org) for the two most commonly implemented open source solutions for IRs.

Neil Jacobs' **Digital repositories in UK universities and colleges** (www.freepint.com/issues/160206.htm) for a recent view from the manager of the JISC Digital Repositories development programme.

The **Loughborough Institutional Repository** https://magpie.lboro.ac.uk/dspace/

Open Archives Forum (http://www.oaforum.org/) for straightforward descriptions of OAI and OAI-PMH.

Alma Swan's **JISC Open Access Briefing Paper** (http://www.jisc.ac.uk/uploaded_documents/ JISC-BP-OpenAccess-v1-final.pdf) for a succinct summary of open access publishing and the role of IRs.

References

Antelman, K. (2004) Do open-access articles have a greater research impact? *College and Research Libraries* [online], 65(5), 372-382. http://www.lib.ncsu.edu/staff/kantelman/do_open_access_CRL.pdf, [accessed 26.05.05].

Crow, R. (2002) The case for institutional repositories: a SPARC position paper. (http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf), [accessed 26.05.05].

Day, M. (2003) *Prospects for institutional e-print repositories in the United Kingdom*. (http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/impact/), [accessed 16.06.05].

EPrints.org (2005) Journal policies – Summary statistics so far. http://romeo.eprints.org/stats.php, [accessed 19.02.06].

Foster, N.F. and Gibbons, S. (2005) Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine* [online], 11(1). (http://www.dlib.org/dlib/january05/foster/01foster.html), [accessed 26.05.05].

Gadd, E. *et al*. (2003a) RoMEO studies 1: The impact of copyright ownership on author-self-archiving. *Journal of Documentation* [online], 59(3), 243-277. (http://iris.emeraldinsight.com/...), [accessed 13.06.05]

Harnad, S. *et al*. (2003) Mandated online RAE CVs linked to university eprint archives: enhancing UK research impact and assessment. *Ariadne* [online], 35(April 2003). (http://www.ariadne.ac.uk/issue35/harnad/), [accessed 19.06.05].

Heery, R. and Anderson, S. (2005) *Digital repositories review*. UKOLN. (http://www.jisc.ac.uk/uploaded_documents/rep-review-final-20050220.pdf), [accessed 30.06.05].

Horwood, L. *et al.* (2004) OAI compliant institutional repositories and the role of library staff. *Library Management* [online], 24(4/5), 170-176. (http://iris.emeraldinsight.com/...), [accessed 26.05.05].

Hubbard, B. (2003) SHERPA and institutional repositories. *Serials* [online], 16(3), 243-247. (http://uksg.metapress.com/...), [accessed 26.06.05].

Kurtz, M.J. (2004) *Restrictive access policies cut readership of electronic research journal articles by a factor of two*. (http://opcit.eprints.org/feb19oa/kurtz.pdf), [accessed 20.06.05].

Pickton, M.J. and McKnight, C. (2006) Research students and the Loughborough institutional repository. *Journal of Librarianship and Information Science*, 38 (4): forthcoming.

Pinfield, S. (2002) Creating institutional e-print repositories. *Serials* [online], 15(3), 261-264. (http://uksg.metapress.com/...), [accessed 16.06.05].

Steele, C. (2003) Digital libraries and repositories. *Information Management Report*, March, 1-5.

Swan, A. *et al*. (2005b) *Delivery, management and access model for e-prints and open access journals within further and higher education*. (http://eprints.ecs.soton.ac.uk/11001/01/E-prints_delivery_model.pdf), [accessed 30.06.05].

Yeates, R. (2003) Institutional repositories. *VINE: The Journal of Information and Knowledge Management Systems* [online], 33(2), 96-100. (http://hermia.emeraldinsight.com/...), [accessed 26.05.05].