Pinfield, Stephen (2002) Creating institutional e-print repositories. Serials, 15 (3). pp. 261-264.

Creating institutional e-print repositories

Stephen Pinfield

Published in: *Serials*, 15, 3, November 2002, pp. 261-264.

This article aims to address three questions:

- What are 'institutional e-print repositories'?
- Why create them?
- How can they be created?

What are institutional e-print repositories?

'E-prints' are electronic copies of research papers or similar research output. They might be 'pre-prints' (pre-refereed papers), 'post-prints' (post-refereed papers), conference papers, book chapters, reports, or related kinds of material. Online collections of such material are sometimes called e-print 'archives'. The term 'archive' is however ambiguous. For many it implies a sophisticated system of the curation and preservation which may not necessarily be in place for e-prints. Some people therefore prefer to use the more neutral 'repository'.

E-print repositories are usually made freely available on the internet. The principle of so-called 'open access' is one supported by many advocates of e-prints[1]. The term 'open' in the context of e-prints does not always refer to access however. The Open Archives Initiative uses the term in a more technical way to indicate system interoperability[2]. The OAI has produced a protocol which allows archives to expose information about their contents on the internet in the form of structured metadata which can be automatically collected up and placed in searchable databases. Any e-print repository which is OAI-compliant can therefore be a part of an interoperable network of databases, the metadata from which can be easily searched by users. Making searching easy helps to ensure that e-prints do not get buried in the chaotic morass of the material on the web. When users have found information about what they are looking for, in an open access environment, they can then link directly to the full text wherever it is located.

There are already a number of successful e-print repositories in existence. These include arXiv, for physics, mathematics and computer science, and CogPrints, for cognitive science[3]. Both of these services are OAI-compliant and make content freely available on the internet. They are centralised subject-based archives, where authors from many institutions mount their e-prints (a process sometimes referred to as 'self-archiving') on a single database held in a single location. But so far most other subject disciplines seem to have shown little inclination to go down the e-print road. To address this situation, many advocates of e-prints suggest that *institutions* should create repositories and at the same time encourage researchers from *all* disciplines to contribute to them. Institutional repositories would contain material produced by members of that particular institution from across the range of subjects. Institutions,

it is argued, have the resources to subsidise archive start-up, technical and organisational infrastructures to support archive maintenance, and also an interest in disseminating archive content[4].

Why institutional e-print repositories?

E-print repositories are a response to a number of structural problems with the current scholarly publishing process. Researchers give away their output in the form of journal articles to publishers with the aim of achieving impact of not income. They want their work to be read, cited and built on by colleagues in their subject field. It is therefore in their interests that their content should be disseminated as widely as possible. However, commercial publishers normally try to restrict dissemination based on subscription. In doing so, they create 'impact barriers' for authors. They also create 'access barriers'. In a world where there are well over 20,000 peer reviewed journals and, most libraries cannot afford subscriptions to even half of these. Therefore, most researchers do not have easy access to most of the literature[5].

In this context e-print repositories create a number of potential benefits. These apply first for the researcher, secondly for the institution, and thirdly for the research community as a whole. For the researcher, e-print repositories have the potential to lower impact and access barriers. They create a situation where content can be disseminated widely and rapidly, it can be easily located and freely accessed. In addition, there is the potential for beneficial spin-offs for the researcher which e-print archives might create. These could include personal hit counts and citation analyses, tools for which are already being developed[6].

For the institution, there are major benefits in a raising its profile and prestige within the research community and beyond. There are also possible practical benefits associated with accreditation and 'information asset' management. The institution becomes aware of and better able to manage research output. In addition, there are potential long-term savings in subscription costs of journals. Freely available content will perhaps mean that some publishers have to scale down their subscription prices and re-focus their activities on managing the peer review process and adding value to the raw content.

All of these benefits add up to benefits for their research community as a whole. E-print repositories have the potential to free-up the research communication process. 'Free-up' in the sense of making it easier and quicker and in the sense of making it cheaper. Better research communication makes things such as unnecessary duplication of research less likely.

Despite these potential benefits, there are a number of common concerns which are raised in relation to creating the e-print repositories. These include issues such as quality control (particularly concerns about the peer review process), intellectual property rights (particularly copyright), the undermining of the 'tried and tested' methods of communication (particularly journals), and the potential increase in workload for staff (particularly in having to self archive their papers). Concerns like these have been addressed in detail by advocates of e-prints, such as Stevan Harnad[7]. One important general point which may be emphasised here, however, is that institutional repositories do not necessarily have to *replace* existing peer-reviewed

journals but might rather *complement* them. The two can exist side by side. Authors should be encouraged where possible to self-archive their e-prints *as well as* publishing them in the peer-reviewed literature.

How can e-print repositories be created?

Initial installation of an OAI-compliant e-print archive is relatively straightforward[8]. Free software is available to do this from e-prints.org[9]. This software provides database technology to organise the e-prints and also a web interface for depositing them and using them.

But setting up any e-print repository is not just a technical matter. It also involves making a number of important collection management policy decisions. These include:

- Document type: will pre-prints be included or only post-prints?
- Document format: HTML, PDF, postscript, etc?
- Digital preservation policies: what will be preserved and how?
- Submission procedures: how will files be formatted and then deposited?
- IPR policies: what are the rights of the author, institution and publisher?
- Metadata quality standards: who will create metadata and what standards and quality thresholds will be applied?

The costs of setting up an e-print repository are in the short term relatively low. Installation involves the cost of a server plus several days of technical staff time to install and configure the software. There are then staff costs in spending time developing policies. However, after initial installation there are a number of more significant costs. There are ongoing costs associated with advocacy (encouraging researchers to submit content), support (helping them to prepare and deposit content), and metadata creation or enhancement (ensuring the content is adequately described). Over time there will also be costs associated with digital preservation.

The real challenges here are the cultural and organisational ones. A great deal of work needs to be done in talking to researchers and encouraging them to think in new ways about disseminating their research output. The model of e-print archives needs testing in different subject disciplines (where conventions of publication may differ from say physics)[10]. The extent to which e-print archives may fit into institutional policies and procedures also needs investigating further in practical ways.

To help kick-start these sorts of investigations, JISC has recently funded a series of projects in this area as part of its FAIR (Focus on Access to Institutional Resources) programme[11]. These projects are currently getting under way and will be worth following over the next two to three years. A number of them (such as the SHERPA project based at the University of Nottingham[12]) concentrate in particular on the development of e-print archives within institutions and it is hoped that they will disseminate the lessons they learn to the wider academic and publishing community.

Conclusion

It remains to be seen whether e-prints in general and institutional e-print repositories in particular will be useful ways of improving scholarly communication. They have potential but this needs testing out. What *is* clear is that the existing system of publishing, which developed in a paper-based world, is looking more and more anomalous and inefficient in the web-based world. We need to try something else.

Stephen Pinfield is Assistant Director of Information Services at The University of Nottingham. He is a member of the CURL (Consortium of University Research Libraries) Task Force for Scholarly Communication. He is also Director of the new JISC-funded SHERPA project which aims to set up institutional e-print repositories in a number of research universities.

---

[1] See the Budapest Open Access Initiative, http://www.soros.org/openaccess/.
[2] The Open Archives Initiative, http://www.openarchives.org.
[3] arXiv, http://www.arxiv.org; CogPrints, http://cogprints.ecs.soton.ac.uk/.
[4] See Raym Crow *The case for institutional repositories: a SPARC position paper*. Washington, DC: SPARC, 2002. Release 1.0. Available at http://www.arl.org/sparc/IR/ir.html.
[5] See Stevan Harnad, 'The self-archiving initiative', *Nature: Debates*. Available at http://www.nature.com/nature/debates/e-access/Articles/harnad.html.
[6] See http://citebase.eprints.org/.
[7] See Stevan Harnad, *For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now*, Section 8. Available at http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm#8.
[8] See Stephen Pinfield, Mike Gardner and John MacColl 'Setting up an institutional e-print archive'. *Ariadne*, 31, March-April 2002. Available at http://www.ariadne.ac.uk/issue31/eprint-archives/.
[9] See http://www.eprints.org.
[10] See Stephen Pinfield 'How do physicists use an e-print archive? Implications for institutional e-print services'. *D-Lib Magazine*, 7, 12, December 2001. Available at http://www.dlib.org/dlib/december01/pinfield/12pinfield.html.
[11] See http://www.jisc.ac.uk/dner/development/programmes/fair.html.
[12] See http://www.sherpa.ac.uk.