OAI compliant institutional repositories and the role of library staff

Lynne Horwood
Shirley Sullivan
Eve Young
Jane Garner
Information Division
University of Melbourne

Biographies:

Lynne Horwood is a librarian in the Teaching, Learning & Research Support Department at University of Melbourne.

Shirley Sullivan is the Electronic Information Coordinator in the Information Division of the University of Melbourne.

Jane Garner is a librarian in the Teaching, Learning & Research Support Department at University of Melbourne.

Eve Young is Metadata Coordinator in the Information Division of the University of Melbourne.

Weaving library expertise throughout the learning, teaching, research, and service functions of the institution. (Lougee, 2003)

Keywords

Eprints, Scholarly communication, Academic libraries, Institutional repositories, Open Archives Initiative, Protocol for Metadata Harvesting

Abstract

The role of librarians in the development and promotion of institutional repositories is discussed. It is presented as a continuation of their existing functions of acquiring, organising and making readily available the resources needed by academic staff and students. Library staff are collaborating with IT staff and academics to disseminate scholarly material and learning objects emanating from their institutions. The Open Archives Initiative and its Protocol for Metadata Harvesting, which provide the technical structure to support the repositories and enable their interoperability for searching purposes, are discussed. The benefits to institutions and their staffs are also reported. The skills needed by library staff are outlined, as well as the pitfalls and problems they may face in persuading academic staff of the virtues of institutional repositories.

INTRODUCTION AND CONTEXT

Institutional librarians are well placed to exercise leadership in information management, based on their print and digital content management expertise. As Herbert Van de Sompel says, "Libraries are there to facilitate access to scholarly information." (Van de Sompel, 2003) Traditionally, academic library staff have acquired, organised, and disseminated scholarly information. The role of librarians is

now expanding to include managing electronic scholarly products and participating in the evolving scholarly communication process.

Libraries not only acquire electronic resources, but also create them. Libraries are being funded to digitise valuable parts of their special collections, both to preserve the originals and to make the content more readily accessible. Notable examples are the work being done by the Oxford Text Archive¹ and, in the Australian context, the SETIS project².

Librarians are increasingly working with academic colleagues to provide online content for research, learning and teaching. Providing access to digital content is an essential prerequisite for institutions establishing and offering flexible online learning delivery.

BACKGROUND

One way of providing and organising digital content is by establishing institutional repositories. Participants at the inaugural meeting in October 1999 of what became known as the Open Archives Initiative <<http://www.openarchives.org/>> were academic librarians and computer scientists interested in archiving, metadata, and interoperability. By the end of October the initiative had changed its name from the Universal Preprint Service (UPS) to the Open Archives Initiative (OAI). This change of name reflected the wider utility expected of the software, which was no longer seen as restricted to eprint repositories. Since that inaugural meeting, there has been growing interest within academia and research institutions in the establishment of repositories to centralise and make readily accessible the intellectual capital of the

¹See <http://ota.ahds.ac.uk/>

² See <http://setis.library.usyd.edu.au/oztexts/>

institution. In some cases, repositories are restricted to research output, but frequently they include teaching and learning materials produced by academic staff.

INSTITUTIONAL REPOSITORIES

In the view of Clifford Lynch, Director of the Coalition for Networked Information (CNI), establishing institutional repositories is a recognition that the intellectual life and scholarship of universities will increasingly be represented, documented, and shared in digital form. (Lynch, 2003) It has been a logical step for the institutional libraries to take the lead in ensuring availability and preservation of research output. In many institutions, such as University of Melbourne, the Australian National University, University of Nottingham, and University of Glasgow, library staff are taking the initiative to establish eprint repositories compliant with international standards. In these universities and many others, the software employed is eprints.org,³ emanating from the University of Southampton.

Another example of institutional libraries leading the way in developing OAI compliant digital repositories is MIT Libraries, where library staff worked with the Hewlett-Packard Company to develop DSpace. This development is an extension of MIT's mission to collect, make available, and preserve important scholarly material of all kinds, especially that of MIT's own academic and research community. The work undertaken by the MIT Libraries reflects the current trends in scholarly communication and education, and offers new means of distributing research material that are enabled by network technology. (Barton, 2003)

Librarians' involvement includes developing content management policies, deciding on what metadata to store and present, crafting author permission and copyright

³ See <http://www.eprints.org/>

agreements, creating document submission instructions, training staff and authors in using the software to submit content, and marketing the repository concept to prospective depositors. (Crow, 2002, p.28) Librarians have the technical skills to assist authors to deposit research material in the repository. They have expertise in dealing with a wide variety of formats. In addition to digital textual resources, librarians are also handling formats such as statistical, mapping, graphical, sound, and moving images. (Pinfield, 2001) In addition, it is the skills and expertise of librarians in the areas of communication, preservation, metadata handling, advocacy and promotion that make them ideal managers of institutional repositories. This article discusses the need for these latter skills further.

METADATA

...interoperability comprises persistent naming, standardized metadata formats, and a metadata harvesting protocol. (Crow, 2002, p.10)

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) defines a mechanism for harvesting XML (Extensible Mark-up Language) formatted metadata from individual repositories. The key component of the interoperability architecture is the use of Unqualified Dublin Core (DC). This comprises 15 elements. Definitions and recommended use of these elements can be found at the Dublin Core Metadata Initiative website <<http://www.dublincore.org>>.

To lower barriers to acceptance, unqualified DC is mandated as a minimum standard for OAI compliant repositories. However, OAI-PMH supports any metadata format that can be described with an XML schema and the use of richer additional metadata formats is encouraged. Greg Simpson and Malcom Polfremann, reporting on Herbert Van de Sompel's keynote address at the 3rd Open Archives Forum in Berlin in March

2003, state that, “the Dublin Core serves as a unifying set of metadata to allow discovery across all communities”, (Simpson, 2003) supporting “metadata packages that are modular, overlapping, extensible, and community-based.” (Simpson, 2003) Herbert Van de Sompel mentioned in this keynote presentation that the intention is to

build upon the OAI-PMH to move beyond interoperability at the level of discovery:

- ❑ references
- ❑ usage logs
- ❑ certification metadata
- ❑ rights metadata (Van de Sompel, 2003)

Metadata without tears

Using eprints.org software, metadata is created automatically as authors enter their documents into the repository. Authors are prompted to enter details such as their name, the title of the document they are entering, and keywords describing the document. The metadata generated in the entering process allows for later retrieval via OAI compliant search engines.

But, “the simplicity of Dublin Core can be both a strength and a weakness.” (National Information Standards Association, 2001) As discussed above, simplicity lowers the cost of creating metadata because it can be created by authors as they self-deposit. But it is crucial to have some reviewing mechanism to ensure the metadata is accurate and sufficiently detailed. Eprint administrators may need to add additional subject headings and keywords. Users cannot be expected to correctly implement a ‘controlled vocabulary’. Librarians are the professionals who have dealt with synonym lists and thesauri to cope with search terms entered being translated into search terms that deliver results. (McLeod, 2000)

The eprints.org software builds the reviewing process into the workflow. An item has to be approved by the system administrator before it goes live. The administrator can accept, edit/enhance or bounce a submission at that stage. (Pinfield 2002) The librarians take on the responsibility for the production of metadata and act as agents for quality control to ensure that records created comply with international standards. (Pinfield 2001) Organising and maintaining digital content, especially that submitted by institutional authors, is a key function of librarians in academic and research institutions.

COMMUNICATION SKILLS

Communication skills are an essential part of formal and informal liaison with academics. In many institutions, librarians are assisting academics to self-deposit their research output into institutional repositories. Good communication and team skills are also needed for librarians working and negotiating with colleagues in the library, IT services and academic departments. (Pinfield, 2001)

ADVOCACY AND PROMOTION

Eprint repositories, by capturing, preserving, and disseminating a university's collective intellectual capital, indicate the quantity and quality of a university's research output. (Crow, 2002) The creation of OAI compliant search engines provides a potential worldwide audience for the institution's research. As Sir John Sulston so cogently stated when accepting the Nobel Prize for Medicine in 2002, "From sharing, discovery is accelerated in the community. Research is hastened when people share results freely." (Meek, 2002) This increased accessibility provided by OAI compliance heightens the impact of the institutional research on the scholarly

community, thus increasing the institution's standing and creating a climate conducive to investment in research by funding bodies.

Librarians employ a variety of methods to publicise and promote institutional open access repositories and demonstrate their value to time stretched academic staff. At the University of Melbourne, senior library staff have visited Deans and Heads of Departments to gather support. The more senior managers can be persuaded to embrace the institutional archiving concept, the more they will encourage colleagues to contribute. (Pinfield, 2001) The most important, but the most difficult, thing is getting content in place. To persuade academic staff to contribute content to the repository, library staff at the University of Melbourne use a number of approaches. They maintain a promotional website, and published an article in the university newspaper. They hold seminars to raise the eprints profile. Attendance at departmental staff meetings provides further opportunities to remind busy academic staff about the value of the repository. Showing the number of hits on individual papers over a period of time demonstrates how open access can increase exposure to research material within shorter timeframes than is possible with the traditional scholarly publishing cycle. In particular, University of Melbourne staff noted that theses placed in the repository receive large numbers of hits from all around the world.

Librarians at the University of Melbourne have worked with academic staff to develop and promote a number of initiatives to assist in the acceptance of the eprint repository by academic authors. Many research papers are already being made available through academics' personal home pages, and the web pages of research projects, laboratories and departments. While this demonstrates a desire for disseminating research output, publishing on such websites is a less effective mechanism than using OAI compliant repositories. As Crow (Crow, 2002, p. 23)

comments, the principal benefit for authors is an enhanced professional visibility due to the increased article impact that open access makes possible. The UMER (University of Melbourne Eprint Repository) website includes instructions on how to link papers in UMER to the website of an academic staff member, so they gain the advantage of world wide access that OAI compliance provides, while simultaneously maintaining a local web presence. Use of a unique author identification number embedded in a UMER URL allows authors to link from their own home/departmental page to their deposited papers in the repository. This time-saving feature means that their home page will always point to the most recently updated list of their papers.

See, for example, the website of Dr Peter Chen at

<http://www.politics.unimelb.edu.au/about/staff/chen.html>.

INTELLECTUAL PROPERTY ISSUES

There are a number of common concerns related to establishing and maintaining an eprint repository that librarians may have to face. These include self-archiving concerns such as the breaking of already established agreements with traditional publishers, the legality of subsequent publication of the research in traditional journals and the possible bypassing of the peer review process. Other concerns are likely to focus on mediation and costs.

Library staff have a role in broadening perspectives of both academic and administrative staff on copyright issues. One academic concern frequently raised is that placing material in an eprint repository precludes its later publication in scholarly journals. A survey by Gadd and colleagues indicates that for those who do not deposit material in institutional repositories their greatest concern was that they would not be published, or that self-archiving would break existing agreements. They were also anxious about the potential loss of integrity of their papers. (Gadd, 2003)

Librarians need to inform academics that depositing eprints in institutional repositories does not necessarily preclude them from publishing in high-impact journals. Many journals now accept items previously available in repositories. To help allay concern, University of Melbourne library staff maintain a list of journals' policies relating to prior publication in repositories.⁴

Library staff can encourage authors to negotiate with publishers so that they retain the right to publish their material in open access repositories as well as in scholarly journals. Some publishers allow authors to retain the copyright for their papers and permit them to deposit material in publicly accessible archives. Others do not.⁵ When a paper has been accepted for publication in a journal, the author assigns the copyright to the publisher or (sometimes) grants an 'exclusive licence' to publish. (Nixon, 2002) But authors need to retain copyright and grant *non*-exclusive licences to publishers. (Crow, 2002 p.21) The response of academic authors themselves to such policies runs the gamut from strict adherence to utter indifference,⁶ but ignorance should not be the basis for agreement to restrictive contracts. (Crow, 2002 p.11) Often authors simply agree to the default terms. Many are surprised to hear that they can edit, adjust or propose language in publisher contracts for their

⁴See Project Romeo <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html>

⁵ For example, the latest copyright agreement issued by Nature Publishing Group asks authors to grant them an exclusive licence to publish. Authors are allowed to "re-use the papers in any printed volume of which they are an author; to post a PDF copy on their own (not-for-profit) website; to copy (and for their institutions to copy) their papers for use in coursework teaching; and to re-use figures and tables" (Nature, 2003). However, the licence expressly excludes "open archival websites, such as those that host collections of articles by an institution's researchers." (Day)

⁶ From a survey of authors and publishers, Gadd, Oppenheim & Probets (2003) found that around a third of academics were not sure who owned the copyrights in a research paper. The same study showed that while 41 per cent of the surveyed academics freely assigned copyright to publishers, almost half (49%) did so reluctantly. As, Bide (2002, p. 24) comments, the "pressure on academic authors to publish (and to publish in high profile journals) may lead them to sign agreements that they may otherwise might not." (Day)

accepted articles. It is a role of library staff, acting as repository administrators, to promote such knowledge.⁷

Cox's (Cox, 2003) study for the Association of Learned and Professional Society Publishers reports the following findings:

- ❑ Large publishers are more likely to allow articles to be posted prior to publication.
- ❑ Most publishers allow published articles to be posted to web sites, though medium-sized commercial companies are an exception.
- ❑ Publishers are more inclined to allow posting published articles to the author's own site than their institution's, but posting to preprint servers is much less acceptable.
- ❑ Few publishers disallow the re-use of authors' material within the academic institution, subject to proper acknowledgement of the journal and publisher.
- ❑ Over half the publishers allow the author to use his or her material within his or her own publications.
- ❑ Although most publishers still require journal authors to assign copyright, the author is not restricted from using the work for personal or institutional purposes.

PEER REVIEW

⁷ The Scholarly Communications Group at Johns Hopkins are encouraging staff to retain certain rights, such as to post their papers on the web, to use them freely in their classes, or deposit them in an institutional repository. JHU actually has a "copyright retention" form that staff are supposed to use when they publish. A website <http://openaccess.jhu.edu/> provides information about publisher's relevant policies.

Another objection to eprint repositories librarians may face is that it might enable the bypassing of the traditional processes of peer-review. Review is an essential part of the existing scholarly publishing process, especially in disciplines like medicine or chemistry. Therefore, librarians need to stress to authors that an institutional repository can include both peer-reviewed and non peer-reviewed items. Many items contributed to eprint repositories have undergone the peer review process. Placing the material on the repository is a way of inviting peer comment as part of a prepublication process prior to submission to a scholarly journal.

MEDIATION

Eprint repositories have been designed to encourage authors to submit their own documents. This works well in established repositories, for example, <<http://www.arxiv.org/>>. The eprints.org software has a self-deposit facility. While the process is straightforward, to establish critical mass, submission can be undertaken by library staff on behalf of their academic colleagues. To assist academic staff wanting to submit their own work, library staff at the University of Melbourne created a set of instructions to clarify the existing online loading procedures.

Library staff may also undertake file formatting and conversion, as many academic users do not have the software to convert a word-processed file into a PDF, which is one of the preferred depositing formats.⁸ The repository can also be configured to accept other formats, such as Word and Powerpoint. It may also be necessary to ensure that submitted HTML is properly formatted and cross-browser compatible. (Nixon, 2002)

⁸ Default formats are ASCII, PDF and HTML.

COSTS

Both the eprints.org and DSpace⁹ software are open source, so the major initial cost that institutions must bear is the purchase of hardware. Staff time in loading and configuring the software must also be factored into the cost equation. The significant *ongoing* cost is staff time in maintaining and publicising the repositories. Attendance at forums (departmental meetings, seminars and workshops) is time consuming. Staff time is also required in negotiating intellectual property rights. The cost of creating additional metadata, particularly that associated with the preservation and administration of eprints (James, 2003), is another aspect to be considered in developing strategic and economic plans for the preservation and usability of resources over time.

PRESERVATION

In the print era, librarians supported the archiving function of scholarly communication by physically maintaining and preserving the printed literature. In Victoria, Australia, for example, the major universities established a cooperative store that provides preservation for research materials. In the digital era, archiving of electronic journals is being undertaken by co-operative action by publishers and librarians, frequently funded by philanthropic trusts.

Although the future is uncertain, eprint repositories are likely to house significant material difficult to obtain elsewhere, as they can comprise the corporate memory of research communities. (James, 2003) In Lynch's view, "as we think about institutional repositories today, there is much less redundancy than we have had in our systems

⁹ <http://www.dspace.org>

of print publication and libraries, so any single institutional failure can cause more damage.” (Lynch, 2003) Institutions supporting repositories may not always be aware of their responsibility to ensure the long-term preservation of content. Even when they are, they may not have the organisational infrastructure or technical knowledge to do this successfully. (Day, 2003) Staff with specific knowledge and experience of practical digital preservation may not be available at all within the institution. (James, 2003) Therefore, where it is concluded that particular collections of eprints should be preserved, it may be that external organisations take over this activity. For example, in Australia, CAVAL (Cooperative Action by Victorian Academic Libraries) may be an appropriate body to provide such a service.

RECOMMENDATIONS FOR FURTHER RESEARCH

Mechanisms to support linkage between items in eprint repositories and primary research data will add value to the eprints and to the primary data. In the immediate future, institutional library staff could undertake this task in collaboration with maintainers of publicly available datasets, such as National Center for Biotechnology Information (NCBI). Greater efforts to link and add value to information generated in the research process, through citation of datasets in preprints or in journal articles and the ability to cross-search and access them in a common research information environment could raise the profile of the datasets. (Lord, 2003 p. 47) When data becomes part of an interoperable network of databases, it may be drawn upon by a larger community of users, who may be unaware that they are drawing upon the resources of many different databases simultaneously, as they can be searched as though they are one large database.

CONCLUSION

As digital resources are now commonplace in academic institutions, their management and accessibility form major responsibilities for librarians. Acquiring, creating, and making available electronic resources are an extension of a library role already well established for print materials. Libraries have traditionally managed the key academic information resources of institutions. In the short term, librarians should be active in installing eprint servers locally and smoothing the path for academics to contribute to them. (Pinfield, 2001) A newer and growing role is management of an institution's intellectual capital, including involvement with other institutional stakeholders in content management systems, virtual learning environments and knowledge management systems.

Librarians involved in institutional repository projects have reported that the effort and organisational costs required to address repository policy, content management, and promotion to academic staff dwarf the technical implementation effort. The challenge for librarians will not be the technical implementation of an eprints service but effecting the cultural change necessary for it to become an integral part of the activities of the institution. (Nixon, 2002) Although the future shape of scholarly communication remains unclear, what is clear is that library and information professionals have key roles to play.

References

- Barton, M. R. & Walker, J. H. (2003), "Building a business plan for Dspace, MIT Libraries' Digital Institutional Repository", *Journal of Digital Information*, Vol. 4, No.2, June. Available <http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Barton/barton-final.pdf>.
- Cox, J. & Cox, L. (2003), *Scholarly Publishing Practice: the ALPSP Report on Academic Journal Publishers' Policies and Practices in Online Publishing. Summary*, John Cox Associates for Association of Learned and Professional Society Publishers, Rockwood, UK. Available <http://www.alpsp.org/news/sppssummary0603.pdf>.

Crow, R. (2002), *The Case for Institutional Repositories: A SPARC Position Paper*. Available <http://www.arl.org/sparc/IR/ir.html>.

Day, M. (2003), *Prospects for institutional eprint repositories in the United Kingdom: ePrints UK supporting study, no. 1, Version 1.0, 28 May 2003*. Available <http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/impact/>.

Gadd, E., Oppenheim, C. & Proberts, S. (2003), *RoMEO Studies 2: How academics want to protect their open-access research papers*, Available <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/>.

Harboe-Ree, C. (2003), *Connecting the Dots: An Integrated Approach to Digital Content Management in an Institutional Context. Technical Report, Monash University Library, Monash University*. Available <http://eprint.monash.edu.au/archive/00000026/>.

James, H., Ruusalepp, R., Anderson, S. & Pinfield, S. (2003), *Feasibility and Requirements Study on Preservation of Eprints. Report Commissioned by the Joint Information Systems Committee (JISC)*. Version 1.0 May 9, 2003. Available http://www.jisc.ac.uk/uploaded_documents/eprints_report_1-0.pdf.

Little, D. (2002), *[Report on] 2nd Open Archives Forum Workshop: "Open access to hidden resources", Lisbon, 5-7 Dec 2002 Breakout Session no 2. Role of Libraries in Open Archives*. Available <http://newsite.david-little.org.uk/metadata/lisbonoai.html>.

Lord, P. & Macdonald, A. (2003), *Data Curation for e-Science in the UK: an Audit to Establish Requirements for Future Curation and Provision The Digital Archiving Consultancy Report commissioned by the JISC Committee for the Support of Research (JCSR) Consultation draft*. Available http://www.jisc.ac.uk/uploaded_documents/e-Sciencefinaldraft.pdf.

Lougee, W.P. (2003), "Diffuse libraries", *ARL Membership Meeting 142 Proceedings*, Available <http://www.arl.org/arl/proceedings/142/lougee.html>.

Lynch, C. (2003), "Institutional repositories: essential infrastructure for scholarship in the digital age", *ARL Bimonthly Report*, 226. Available <http://www.arl.org/newsltr/226/ir.html>.

McLeod, B. (2000), "Factoring the librarian back into world wide web cataloguing", *AusWeb2K-The Sixth Australian World Wide Web Conference*, Cairns, Australia. Available <http://ausweb.scu.edu.au/aw2k/papers/mcleod/paper.html>.

Meek, J. (2002), "Nobel prize for British and US scientists who used worms to decode the book of life." *The Guardian*, 8 October 2002. Available <http://www.guardian.co.uk/Print/0,3858,4517292,00.html>.

National Information Standards Organization (2001), *Dublin Core Metadata Element Set, ANSI/NISO Z39.85-2001*. Available <http://www.niso.org/standards/resources/Z39-85.pdf>.

Nixon, W. (2002), "The evolution of an institutional eprints archive at the University of Glasgow", *Ariadne*, Issue 32, June-July. Available <http://www.ariadne.ac.uk/issue32/eprint-archives/intro.html>.

Pinfield, S. (2001), "Managing electronic library services: current issues in UK higher education institutions", *Ariadne*, Issue 29, October. Available <http://www.ariadne.ac.uk/issue29/pinfield/intro.html>.

Pinfield, S., Gardner, M. & MacColl, J. (2002), "Setting up an institutional eprint archive", *Ariadne*, Issue 31, April. Available <http://www.ariadne.ac.uk/issue31/eprint-archives/intro.html>.

Simpson, G. & Polfremann, M. (2003), *Report from 'Networking Multimedia Resources', the 3rd Open Archives Forum, held in Berlin, 27-29th March 2003*. http://ahds.ac.uk/third_open_archives_forum_report.pdf.

Smith, M. (2002), "DSpace: An Institutional Repository from the MIT Libraries and Hewlett Packard Laboratories", *Research and Advanced Technology for Digital Technology 6th European Conference, ECDL 2002, Rome, Italy*, Springer, Berlin, pp. 543-549.

Van de Sompel, H. (2003), "The OAI and OAI-PMH: how did we get here, and where do we go from here?" *3rd Open Archives Forum*, Berlin. Available http://www.oaforum.org/workshops/berl_programme.php.

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Horwood, L;Sullivan, S;Young, E;Garner, J

Title:

OAI compliant institutional repositories and the role of library staff

Date:

2004-05-01

Citation:

Horwood, L., Sullivan, S., Young, E. & Garner, J. (2004). OAI compliant institutional repositories and the role of library staff. Library Management, 25 (4/5), pp.170-176. <https://doi.org/10.1108/01435120410533756>.

Publication Status:

Published

Persistent Link:

<http://hdl.handle.net/11343/33742>