

All Change : the Ever Evolving Institutional Repository at the University of Melbourne

Introduction

Much has already been written on the evolution and growth of open access institutional repositories (IRs) and it is not the purpose of this paper to recap developments. There are a few seminal works that serve as excellent introductions to the topic, chief among them are those of Raym Crow (Crow 2002) and Clifford Lynch (Lynch 2003). Since these articles were published, much has been written and much has changed, as more institutions have become involved in creating repositories. There are sites which provide statistics on the existence and growth of IRs, for example, ROAR, the Registry of Open Access Repositories. (<<http://roar.eprints.org/>>) This paper will provide a brief background on the history and growth of institutional repositories, and will concentrate on the institutional repository at the University of Melbourne, and the workflow implications for staff in the Information Access program within the Information Services Division.

Institutional Repositories

As Raym Crow states, institutional repositories are, “digital collections capturing and preserving the intellectual output of a single or multi-university community” (Crow 2002). Clifford Lynch’s definition of an institutional repository is that it “is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. Services as well as content form the basis of a repository.” (Lynch 2003)

As Pringle describes them, “repositories represent new ways of organising research and are taking shape in a variety of experimental forms. They vary in the types of content, the purposes of their creators, and their relationship to researchers.” (Pringle 2005) Different institutional repositories have different goals. An educational or research institution may use them to collect and showcase their research or teaching materials. A government repository’s purpose may be to disseminate information to the general public.

IRs are being established for a variety of reasons: to provide open access to the institution’s output, as a showcase for research outputs, to enhance learning outcomes through sharing and re-use of high quality learning materials, to support preservation, or to prepare for the Research Quality Framework (RQF), or its equivalents in other countries.

Content accepted into IRs varies according to each institution’s collection policy. They may incorporate any work produced by the institution’s academic staff, students or professional staff. The material might include, for example, not only theses but students’ honours projects. Content generally includes research articles, conference papers, technical reports, working papers, books and book chapters. They may include learning

objects such as lecture notes and other course materials, but most exclude administrative data.

Background and history

IRs are mostly being established using Open Archives Initiative (OAI) compliant software. OAI is an international movement that encourages data sharing by developing and promoting technical standards so that distributed repositories form part of a global network of cross-searchable repositories, increasing the visibility, accessibility and impact of their contents.

The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. OAI has its roots in the open access and institutional repository movements. Continued support of this work remains a cornerstone of the Open Archives program. Over time, however, the work of OAI has expanded to promote broad access to digital resources for eScholarship, eLearning, and eScience. (<<http://www.openarchives.org/>>)

An important feature of the OAI is the Protocol for Metadata Harvesting (PMH) which enables the cross searching of all OAI compliant repositories. See URL <<http://www.openarchives.org/OAI/openarchivesprotocol.html>> for further details. Search engines such as OAIster and Google provide the means of searching across all OAI compliant repositories as though they are all one large file. Commercial tools such as Scopus and Web Citation Index are also appearing as tools to search IRs and provide value added services such as citation counts. University of Melbourne IR staff are involved as a data provider in these services, creating links to the global community.

Growth of Open Access

Institutional repositories form a key component of the open access movement to bring scholarly research onto the open web (Quint 2006).

Institutions are collaborating and creating repositories in partnership. The ARROW project (Australian Research Repositories Online to the World) is an example of such a partnership. This is a consortium of institutions which have individual institutional repositories but collaborate with one another to “identify and test software to support best-practice institutional digital repositories at the ARROW Consortium member sites to manage e-prints, digital theses and electronic publishing” and also to “develop and test a national resource discovery service using metadata harvested from the institutional repositories by the National Library of Australia”. (<<http://arrow.edu.au/about/>>).

The ARROW project and APSR (Australian Partnership for Sustainable Repositories, <<http://www.apsr.edu.au/>>) are two projects funded by the Australian government in recognition of the potential value of IRs. The ARC (Australian Research Council) is encouraging the researchers it funds to make the results of funded research available in an

open access repository (ARC 2007). The National Health and Medical Research Council (NHMRC) announced

To maximise the benefits from research, findings need to be disseminated as broadly as possible to allow access by other researchers and the wider community. The NHMRC encourages researchers to consider the benefits of depositing their data and any publications arising from a research project in an appropriate subject and/or institutional repository wherever such a repository is available to the researcher(s). (NHMRC 2006)

This echoes support from other research funders. Scholarly societies, academic institutions, research institutes, etc all see benefits in open access and more and more are requiring deposit in IRs of research papers published in journals. “The Wellcome Trust has taken a lead in its commitment that communication of research results is an essential part of the research process.” (Rightscom 2007)

Benefits of open access

There are two models of open access publishing, open access journals and open access through repositories such as UMER (University of Melbourne ePrints Repository). Open access publication models for scholarly communication enhance the dissemination of research findings to all potential users, increasing the economic and social returns to public investment in research and development (R&D). (Houghton 2006).

Under the current system of scholarly communication, much of the intellectual output and value of an institution’s intellectual property is diffused through thousands of scholarly journals. While faculty publication in these journals reflects positively on the host university, an institutional repository concentrates the intellectual product created by a university’s researchers, making it easier to demonstrate its scientific, social and financial value. (Crow 2002)

“Not only does society as a whole benefit from open access through more effective access to information and an expanded and accelerated research cycle, but the visibility, usage and impact of the work of individual researchers increases”. (Richardson 2006) Studies consistently show the positive link between citation impact and open access. (Houghton 2006) Evidence is beginning to emerge that work that is freely available is in fact cited more than material restricted by fee access. Bergstrom and Lavaty refer to a number of articles which demonstrate the increased citation rate of open access articles compared to those hidden behind subscription barriers. Articles on open access receive at least 50% more citations than those that are subscription based. Open access articles are cited significantly earlier and more frequently than those that are not open access. (Bergstrom 2007)

In a climate of increasing assessment of research quality and impact, an IR raises the profile of the institution by ensuring that the research output of the institution is widely

disseminated and not restricted to universities that can afford subscriptions to expensive scholarly journals. This visibility helps to enhance the institution's reputation and thus its ability to attract high quality researchers and research funds. Additionally the repository provides a central location for storing research output and accompanying metadata, thus easing administrative burdens. Rapid dissemination of results in many areas of research is important to establish priority in a field and to elicit timely feedback.

Since any individual researcher could 'publish' material online, some academics are concerned that self-archiving could lead to 'vanity publishing' of works that have not been subject to quality control; however, papers which are loaded on an institutional repository are first placed in a submission buffer, where they can be scrutinised, before acceptance or rejection. Once the paper has been moved from the buffer to the repository, the work is open for review by experts anywhere in the world.

Benefits for the University of Melbourne

University of Melbourne scholars share the benefits enumerated above.

It is common for academic departments in many disciplines to maintain departmental servers of working and research papers. Many academic staff question why they should add their work to an institutional repository when their papers are already available via a departmental website. Institutional repositories build on and improve the practice of posting work on departmental sites. The main drawback to departmental servers is the lack of ability to search across a range of institutions and departments. The increased exposure of items in the University's IR will improve the opportunity for impact of University of Melbourne researchers in the international academic arena. The benefits for the University include:

- Ability to see which are the most viewed papers and what sites are accessing the papers This will give academics, their departments, and the University administration data on the impact that University of Melbourne research has upon international colleagues.
- The software allows cloning of an existing paper so that revised versions are less time consuming to load.
- The software provides the facility to track revisions or different versions of papers (<http://www.lib.unimelb.edu.au/eprints/home.htm#aim>)

One academic staff member at the University raised a proposal to link lists of publications from UMER to individual and departmental web sites to maintain their research profiles. This has been acted on. The UMER staff members have become closely involved with research groups in assisting them in various trials to help them meet their needs.

UMER

The University of Melbourne established its own repository in 2002. This originally was for the research output of the University's academics but was extended to postgraduate

student research output and also that of professional staff. The collection policy limits ingest to research output rather than including classroom materials and undergraduate student input. Currently items included are pre-prints, working papers, post-prints, journal articles, theses, books, book chapters and research reports.

There is also the possibility of UMER being used by the Melbourne Research and Innovation Office as a record of research output of the university. UMER could be used to assist in the compliance of our reporting requirements to the Government.

Melbourne University's PhD and Masters Research Theses that are in UMER are also accessible via the Australasian Digital Theses database (ADT). Originally, the ADT program hosted our theses in its repository, but now harvests from UMER by using the metadata in the records. It is important to accurately identify the thesis type. We shall go through the process of creating a record for the repository in more detail later on in this paper.

The driving force behind the growth of UMER is the University's *Growing Esteem* agenda. "Growing Esteem reaffirms Melbourne's intention to be one of the finest universities in the world, highly regarded for making distinctive contributions to society in research, learning and teaching and knowledge transfer." (University of Melbourne, Office of the Vice-Chancellor 2007) The following statement in the University's 2006 *Growing Esteem Strategic Plan* highlights the importance of open access publishing.

In addition to addressing research quality, the University will aim to increase research citations by making Melbourne research more easily accessible to other researchers. 'Open Access' publishing will be strongly encouraged, including the establishment of institutional digital repositories of scholarly works. (University of Melbourne 2006)

Closely aligned with this is the Information Services mission.

The Information Services mission is to maximise our value by linking with the University of Melbourne's Growing Esteem strategy and its triple helix through leadership, innovation and quality in information services, systems and technologies, namely, libraries; cultural collections and archives; e-learning services; e-research services; and corporate information management. (University of Melbourne, Information Services 2007)

Another factor which is pushing this work to the fore is the RQF. As the University has to be prepared for the RQF, many trials are being conducted to test the best methods of inputting records into UMER or some other University of Melbourne repository.

Transition

Times have changed. Gone are the days when the word 'library' referred to predominantly storing and lending physical formats such as monographs, journals, microfiche and audiovisual materials. As we rush into the digital era, academic libraries are becoming much more involved in the provision of digital services to the academic community. Libraries are now taking a major role in collecting, managing, preserving and providing access to information in digital formats. Along with these developments have come changes to our workflow.

The traditional method of managing library information, as stated by Tyler O. Walters, is to "select, acquire, organize, make accessible, promote, preserve, and instruct people about how to use these information resources." (Walters 2007) The same can be said when managing the content of, and creating the records for, an institutional repository. The procedures for achieving this end are quite different. Workplace interactions, changing roles and a myriad of new training and knowledge are needed to complete these tasks. As members of a 'technical services' team who have been heavily involved in copy cataloguing and editing of bibliographic records, we have seized opportunities to transition across to the highly regarded and evergrowing field of cataloguing digital objects (metadata creation). The management and dissemination of digital collections has become increasingly important, not only in providing access to the researchers and the wider community, but also in fulfilling the University's and Government's requirements. Even though UMER has existed for 5+ years, self-submission, on the whole, has not been embraced fully by the University's scholarly community. This is being addressed by the Co-ordinator of Digital Repositories, who is marketing and promoting the repository. Through various trials completed by the UMER team, the value of inputting their work has been demonstrated to academics and the result has been that a growing number of faculties are self-depositing.

Workflow

From the Information Access Program a number of people have become involved in the repository work as a part of their day to day work, as UMER work has been absorbed into the existing staffing budget. The workload varies according to deadlines and priorities for different activities. At times, it is only a little, other times a sizeable amount of time; for example, full time for several weeks to meet requirements of special projects such as preparation for the RQF.

The UMER team consists of the Co-ordinator of Digital Repositories, 3 Metadata Group staff members who create the metadata records, and 3 Editorial Review staff who are responsible for quality checking and copyright issues. We, the UMER team, work under the leadership of the Co-ordinator of Digital Repositories. The co-ordinator has the responsibility of liaising with the faculties, research departments, academics and students and other organisational units. She then collaborates with the team to organise the work and direct the workflow to ensure that the collection objectives and metadata standards are met.

Other departments with which the UMER team has direct contact include:

- Special Collections which, as the stewards of the University of Melbourne theses, deal with borrowing and copying requests from both individuals and institutions. When permission is sought from the author for a copy of a thesis to be made, in compliance with the Copyright Act, the author is also asked to grant permission for her/his thesis to be added to UMER and the ADT. These theses are then scanned by the Digitisation Department, a unit of Special Collections, and are then passed on to the UMER team who create the record.
- The RQF Support Team from the Melbourne Research Office works closely with the research community and collaborates with us. They will be responsible for advising the University on any RQF requirements on electronic repositories.
- Record Services staff have helped in scanning special documents as required, for example, where items need special treatment. This may be because they are in colour or the font used caused problems.
- The Themis Team in the Research Office (Themis is the name of the University's Enterprise Management System). UMER staff created links from UMER to the relevant staff profiles in the Find an Experts site on Themis. (<<http://www.findanexpert.unimelb.edu.au/>>)
- Faculties themselves, using research assistants or professional staff, are involved in adding journal articles, conference papers, working papers and research reports to UMER. Geomatics and Economics & Commerce are two faculties that self deposit. UMER staff are available to assist if required.
- Research assistants in the Faculty of Land and Food Resources have already begun to deposit theses into UMER. Editorial staff in UMER are involved in quality checking of the metadata and moving the records from the buffer to open access.

We also deposit theses and departmental publications for the University community. This is done by creating metadata records using ePrint.org open source software. At present, we are concentrating on works that have been requested by the wider community. It is planned that highly sought after theses will be selected for inclusion in UMER. For example, when a copy of a thesis is requested by an institution or an individual, permission must be given by the author. A permission form is sent to the author by Special Collections staff. If permission is granted, the thesis is scanned and is also saved in PDF format. Most authors also grant permission for their theses to be included in the University of Melbourne ePrint Repository and the Australasian Digital Theses Database (ADT).

Record Creation

As with any record creation, we first check that the item has not yet been deposited in UMER.

The software provides different templates for different item types. The ones most commonly used are: journal (paginated), conference paper or thesis. Each template has different mandatory fields that are indicated by an asterisk*. In the case of a thesis, the mandatory fields are status, author or authors, title, subject (which is the faculty or department), year, abstract, thesis type, the department and the institution. Author

name/names are transcribed as appears on the document. Separate fields are used for the given name and the family name, for example, R. H. Smith or Anne Robertson. Only the first letter of the title is capitalised. If there is subtitle it is added after a colon. A full stop is not added as the software adds this. (If someone adds a full stop then the title appears with two full stops.)

We copy and paste the abstract from the thesis. If we receive a document in Word format, this is converted to PDF. In some cases, the item comes to us in several parts or chapters, and these all need to be linked and made as one document before uploading to UMER. We need to take extra care to check the document for correct pagination as formatting can create problems. We use Adobe Acrobat Professional to perform these tasks. It is important to correctly identify the thesis type, as the Australasian Digital Thesis database harvests Research Higher Degree theses using the metadata created in UMER. In UMER, we are able to use Library of Congress subject headings as well as natural language keywords.

We choose the format and then the security level which allows access to the full record or restricts access to the metadata only. The document is then uploaded either by selecting a file from the PC or by capturing from the document's URL. After these fields have been filled in, a metadata record, which is similar to a bibliographic record, has been created.

Items when uploaded go into the buffer and are handled by the Editorial Review staff. Quality issues are dealt with at this stage. The following list indicates the type of checks that need to be made:

- ensuring that self depositors are authorised to submit their work
- proof reading for typographical errors and erratic layout
- making sure that the full text document is loaded successfully. (Young 2003)

The other major aspect of this work is checking for possible copyright issues. Copyright status is checked on the author's behalf using the Sherpa/Romeo publisher copyright policies and self-archiving site (<<http://www.sherpa.ac.uk/romeo.php>>). Staff also clarify the "conditions or restrictions imposed by the publisher which govern archiving rights or activities." (Sherpa/Romeo 2006) Different publishers do have different policies for pre-prints, post-prints and publishers' versions. Copyright issues that cannot be resolved are sent to the Copyright Office.

According to the Sherpa/Romeo service, 138 publishers have agreed that authors can deposit the published version and a further 19 that, although authors cannot deposit the published version, they can deposit a refereed preprint. (Righthstcom 2007)

When the item is ready for open access, it is deposited and a persistent URL is given to the author. This URL can be used on their websites to link to their works in UMER. Authors are also informed about how to view the statistics for their works.

Training and Skills

The complex nature of this work has brought up the need for specialised skills and training as well as some background information for participating staff. In order to contribute to the University community, we have already:

- had general training for metadata record creation using eprint.org software
- had some PDF training using Adobe Professional. This is used to edit items, for example, copying, pasting, making amendments, linking pages, converting Word to PDF
- had regular meetings with UMER group to discuss updates, changes and queries
- attended seminars and forums to gain more background knowledge of issues related to copyright, RQF and *Growing Esteem* and its implications.

Future training in metadata and html/XML is in the pipeline.

Even though the responsibility of sending permission letters to the authors lies with the Special Collections staff, we at times need to directly communicate with these authors to clarify issues related to consent and regarding the format in which their works are created. They are very grateful and appreciative when they view the end result online and can see the value of the statistics page, which shows not only the number of downloads but also the countries from which users have accessed their work.

Future developments

- It has been mandated by the School of Graduate Studies that students starting their thesis by research in 2007 will be required to deposit a digital copy as well as print upon completion.
- Special Collections staff are monitoring requests for print theses. Where there is considerable demand, permission will be requested from the authors and these will be digitised and entered into UMER.
- The software platform underlying UMER may move from the open source ePrints.org software to a commercial product, DigiTool. Because of this potential migration, some cleanup of the UMER records will be needed and every iteration of the new repository will need to be checked against UMER records to ensure all required elements are included.

Conclusion

In order to produce high-quality research, researchers must have easy and rapid access to as wide a range as possible of the data and information produced by other researchers. Similarly, successful dissemination and exploitation of research depends on effective flows of information between researchers and other individuals and organisations with an interest in its results. (Rightscom 2007)

IRs are helping academic librarians reinvent themselves. Library departments are restructuring and pursuing collaborative partners to manage their university's digital

intellectual output. This involves many skills, ranging from the ability to create metadata and integrate information systems to the ability to promote the IR as an indispensable information management tool. (Walters 2007)

We now see ourselves as working more closely to meet the University of Melbourne's "Growing Esteem" agenda. We can see where we fit in the bigger picture. We are working with real people not just names in a journal or monograph. We are being personally thanked for making access available by authors who are excited about their statistics. We are involved in building relationships with the university community. We are involved in Knowledge Transfer which, in turn, leads to increased accessibility and citation impact.

The constantly changing and interesting environment, in which we are working, has encouraged us to remain flexible and adaptable to the changes. Time management and prioritisation are essential attributes. Commitment and continuous learning are the keys that lead to gains in professional development.

References

- ARROW, Australian Research Repositories to the World, accessed 22 April 2007, <<http://arrow.edu.au/about/>>.
- Australian Research Council 2006, *Discovery Projects: Funding Rules for Funding Commencing in 2008*, accessed 22 April 2007, <http://www.arc.gov.au/pdf/DP08_FundingRules.pdf>.
- Bergstrom, T and Lavaty, R 2007, *How often do economists self-archive?* accessed 20 February 2007, <<http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1203&context=ucsbecon>>.
- Crow, R 2002, *The Case for Institutional Repositories, SPARC Position Paper*, accessed 26 March 2007, <http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>.
- Houghton, J, Steele, C and Sheehan, P 2006, *Research Communication Costs in Australia: Emerging Opportunities and Benefits: a Report to the Department of Education, Science and Training*, Canberra, DEST, accessed 5 May 2007, <http://www.dest.gov.au/NR/rdonlyres/0ACB271F-EA7D-4FAF-B3F7-0381F441B175/13935/DEST_Research_Communications_Cost_Report_Sept2006.pdf>.
- Lynch, C 2003, 'Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age' *portal: Libraries and the Academy*, v.3, issue 2, accessed 22 April 2007, <http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v003/3.2lynch.html>.
- National Health and Medical Research Council (Australia) 2006, *Project Grants Funding Policy for Funding Commencing in 2008*, accessed 15 May 2007, <http://www.nhmrc.gov.au/funding/_files/profundingpol.pdf>
- Open Archives Initiative, accessed 19 April 2007, <<http://www.openarchives.org/>>
- Pringle, J 2005, 'Partnering helps institutional repositories thrive', *KnowledgeLink Newsletter*, February, 2005, accessed 14 May 2007, <<http://scientific.thomson.com/news/newsletter/2005-02/8264025/>>.
- SHERPA/RoMEO 2007, '*Publisher Copyright Policies & Self-archiving*', accessed 9 May 2007, <<http://www.sherpa.ac.uk/romeo.php>>.
- Quint, B 2006, 'Institutional Repositories on Target: ARL Survey and Scopus/Scirus Features', *Information Today*, October 5, accessed 7 May 2007, <<http://www.infotoday.com/newsbreaks/nb060828-2.shtml>>.
- Rightscom 2007, *Research Funders' Policies for the Management of Information Outputs A report commissioned by the Research Information Network*, accessed 26 April

2007, <<http://www.rin.ac.uk/files/Funders'%20Policy%20&%20Practice%20-%20Final%20Report.pdf>>.

Richardson, J 2006, 'Research Quality Framework as a catalyst for open access', *AusWeb06: the Twelfth Australasian World Wide Web Conference, Noosa Lakes, Qld, 1-5 July 2006*, accessed 5 May 2007, <<http://ausweb.scu.edu.au/aw06/papers/refereed/richardson/paper.html>>.

University of Melbourne 2006, *Growing Esteem: the University of Melbourne Strategic Plan 2006*, accessed 5 May 2007, <http://www.unimelb.edu.au/publications/docs/strategic_plan2006.pdf>.

University of Melbourne 2007, Find an Expert, accessed 15 May 2007, <<http://www.findanexpert.unimelb.edu.au/>>.

University of Melbourne, Information Services 2007, *Information Services: mission, purpose and values*, accessed 14 May 2007, <<http://www.infodiv.unimelb.edu.au/about/vision.html>>.

University of Melbourne 2006, *Growing Esteem: the University of Melbourne Strategic Plan 2006*, accessed 14 May 2007, <http://www.unimelb.edu.au/publications/docs/strategic_plan2006.pdf>.

University of Melbourne, Office of the Vice-Chancellor 2007, *Growing Esteem*, accessed 14 May 2007, <<http://growingesteem.unimelb.edu.au/>>.

Walters, T. 2007, 'Reinventing the library: how repositories are causing librarians to rethink their professional roles', *portal: Libraries and the Academy*, v.7, no.2, pp 213-225, accessed 1 May 2007, <http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v007/7.2walters.html>.

Young, E & Sullivan, S 2003, 'Establishing an eprint repository at the University of Melbourne: implementation aspects', *Proceedings Educause in Australasia 03*, pp 25-35, viewed 21 April 2007, from <<http://eprints.infodiv.unimelb.edu.au/archive/00000265/02/eveandshirley.pdf>>