Scholarly information repository services at Monash University

Suzanne Clarke (Suzanne.Clarke@lib.monash.edu.au)
Andrew Harrison (Andrew.Harrison@lib.monash.edu.au)
Sam Searle (Samantha.Searle@lib.monash.edu.au)
Monash University, Australia

Abstract

Monash University has a serious commitment to repositories and research data management. This is demonstrated by its leadership role in the National Systemic Information Infrastructure projects ARROW (http://arrow.edu.au/), DART (http://dart.edu.au/) and ARCHER (http://archer.edu.au/) as well as the Monash University Information Management Strategy and specific data management activities. In recognition of its commitment to data management, Monash University is now the lead agent in the Australian National Data Service (ANDS) (http://ands.org.au/). This is a national project which is working to develop good data management practice, collaboration, infrastructure and services.

This paper outlines the development of government policy and programs in Australia as they relate to open access and the development of institutional repositories. It then provides an overview of the Monash University ARROW Repository (http://arrow.monash.edu.au/vital/access/manager/Index). The main content of the paper is an examination of the way in which Monash University has responded to policy drivers around open access by establishing a complex, multi-purpose institutional repository that contains not only research publications, but also research data and multimedia research collections, providing discipline-specific support for publishing and dissemination. Conclusions are made about the future of the ARROW repositories and the learnings from the journey outlined above.

Keywords

Repositories, Data Management, Research, Open Access

1. Introduction

This paper outlines the development of government policy and programs in Australia as they relate to the development of institutional repositories and open access to publically funded research. The paper then considers the national ARROW (Australian Research Repositories Online to the World) project, a partnership of several institutions led by Monash University and operating from 2004-2008. The main content of the paper is an examination of the way in which Monash University has dealt with the various dilemmas involved in establishing a complex, multi-purpose institutional repository, including the development of its role into the management of research data and research publications.

Monash University’s approach to open access is considered in relation to the various categories of content. Monash University has concentrated its efforts on using the repository to bring into the public domain recent publications of Monash University’s research collection and related material that would not otherwise be publically available online, including research publications, research data, research collections and research publication services. Open access dissemination is outlined, including subject specific disciplines. The Library’s DARE Project is considered in terms both of populating the repository and preparing liaison librarians for future roles.

2. Open access public policy development in Australia

The Australian Federal Government has funded a number of initiatives to support programs of e-research, open access, repository development and data management in Australian universities. It has been supportive of the move to open access for publicly funded research.

A number of national initiatives and strategic government papers have identified issues that have and will continue to have an impact upon the role of libraries and librarians. The major initiatives are outlined below:

2.1. eResearch

In 2004 the National Collaborative Research Infrastructure Strategy (NCRIS) was established as part of the program Backing Australia’s Ability – Building our Future through Science and Innovation. The
National e-Research Coordinating Committee identified the discovery and re-use of publicly funded research data as one of the most pressing needs facing the development of research, especially e-research, in Australia. NCRIS is providing $542 million over 2005-2011 to supply researchers with major research facilities supporting infrastructure and networks necessary for world-class research (DIISR 2009b).

A part of NCRIS is the Platforms for Collaboration (PIC). These are programs which are to develop rapidly, creating an ongoing flow of opportunities to enhance the quantity, quality and productivity of research effort. The Australian National Data Service (ANDS), of which Monash University is the lead institution, is one of the PIC investment activities (DEST 2007).

2.2 Data Management

In 2007 the Australian Code for the Responsible Conduct of Research was released to guide institutions and researchers in responsible research practices. The Code (NHMRC 2007) replaces the 1997 joint National Health and Medical Research Council / Australian Vice Chancellors’ Committee Statement and Guidelines on Research Practice. Among other things, the Code outlines requirements for good management of research data and primary materials. The Code states that “Policies are required that address the ownership of research materials and data, their storage, their retention beyond the end of the project, and appropriate access to them by the research community” (NHMRC 2007, sec 2). More specifically, the Code suggests that “Research data should be made available for use by other researchers unless this is prevented by ethical, privacy or confidentiality matters.” (NHMRC 2007, sec 2.5.2).

2.3 Open Access

The Australian Department of Education, Science and Training Accessibility Framework (DEST 2004) aims to ensure that information about publicly funded research and its outcomes is accessible both to end users and to the publicly funded research community. This applies to research data, research publications and other research outputs provided with the aid of public funds (Cook 2008).

The Cutler review Ventourousaustralia: building strength in innovation, released in January 2008, recommends that “Australian governments should adopt international standards of open publishing as far as possible. Material released for public information by Australian governments should be released under a creative commons licence” (Cutler 2008, p.95).

Repositories of research publications and other outputs were required by the Research Quality Framework (RQF) program and now the Excellence in Research for Australia (ERA) program, which is a key part of the new research quality and evaluation system announced recently by the Minister for Innovation, Industry, Science and Research, the Hon. Senator Kim Carr (DIISR 2008). The ERA initiative will be developed by the Australian Research Council (ARC) to assess research quality within Australia's higher education institutions using a combination of indicators and expert review by committees comprising experienced, internationally-recognised experts (ARC 2008b).

2.4 Repositories

The Commonwealth’s Australian Scheme for Higher Education Repositories (ASHER) program (2007-2009) has provided funding to assist with the process of establishing and populating repositories with research outputs, stipulating that the content is to be available on open access where possible. (DIISR 2009a). Government funding agencies are also promoting open access and the development of repositories through changes to funding agreements. ARC Discovery Projects funding rules for funding commencing in 2010 encourage 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”; researchers who do not intend to deposit in a repository within six months must provide an explanation in the project's Final Report (ARC 2008a, p. 36-37). A similar approach is taken by the NHMRC Project Grants Funding Policy for research starting from 2010 (NHMRC 2009).

3. ARROW (Australian Research Repositories Online to the World)

The ARROW project, which concluded in December 2008, was funded by the Australian Government Department of Education, Science and Technology (DEST) under the Research Information Infrastructure Framework for Australian Higher Education, sponsored as part of the Commonwealth Government's Backing Australia's Ability program. Monash University was the lead member of the ARROW Consortium, which included the University of New South Wales, Swinburne University of Technology and the National Library of Australia. The ARROW project identified and tested software
or solutions to support best practice institutional digital repositories comprising e-prints, electronic theses, e-research and electronic publishing.

After developing repository software, the ARROW project broadened its scope to include a range of Australian universities who had adopted this solution. By the end of the project, members of the “ARROW Community” included the Central Queensland University, La Trobe University, Macquarie University, Murdoch University, Queensland University of Technology, University of Ballarat, University of Newcastle, University of New England, University of South Australia, University of Southern Queensland, University of the Sunshine Coast and the University of Western Sydney.

In addition to leading the consortium Monash also established a local repository, the Monash University ARROW Repository, which will now be considered.

4. Monash University ARROW Repository

Monash University, the largest university in Australia, serves approximately 55,000 students (38,500 EFTSU) and 10,000 staff (7,000 EFT) over six Australian and two overseas campuses in Malaysia and South Africa. Monash University has an extensive international and distance education commitment, with more than 12,000 students (8,000 RFT) being external, distributed around the world. Monash University has a tradition as an early adopter of new technologies, and is recognised in Australia as being at the forefront of the digital revolution. Its current strategic direction is focussing on achieving excellence through a cross-disciplinary, multi-campus, international approach. Consistent with this culture the university has invested strongly in information technology to support research, teaching and administration. Over the past five years it has been the lead institution in a number of projects designed to create digital solutions to assist teaching, learning and research.

As a key element within Monash University’s approach to research data storage and access, it is expected that the Monash University ARROW Repository http://arrow.monash.edu.au will provide long term benefits for the University’s research community by:

• creating a publicly accessible central collection of research;
• increasing the visibility and usage of research;
• increasing access to existing research data and materials not available elsewhere; and
• providing a search mechanism to locate specific research outputs, by creator, subject, keyword, or Monash University faculty.

The Monash University ARROW Repository has a policy objective of making as much material open access as possible and currently 100% of material in the repository has a full text document or image available for open access. Monash University has concentrated its efforts on using the repository to bring into the public domain recent publications of Monash University’s research collection and related material that would not otherwise be publically available.

Authors of pre-prints and other unpublished works are encouraged to assign Creative Commons licences to their work, allowing the repository to share their work in an open access format. For this purpose Monash University is using the Attribution-Noncommercial-Share=Alike=2.5-Australia (Creative Commons) model (Creative Commons Organisation 2009).

• The current and planned content of the repository may be summarised as follows:
• Research publication and outputs
• Research data
• Collection of resources supporting collections, including images, newspapers and patents
• Research publishing.

The various categories are considered below.

4.1 Research publication and outputs

The Excellence in Research for Australia (ERA) initiative will assess research quality within Australian higher education institutions in order to identify and promote excellence in research activity in Australia. Evaluation of research will be undertaken in eight discipline clusters using a combination of indictors and expert review by committees comprising internationally recognised experts. (ARC, 2009c)

The evaluation process of the ERA, will require research outputs to be made available to reviewers. These research outputs will be loaded into the Monash University institutional repository. The accessibility of the research outputs will depend upon nature of the material and ownership of intellectual content in relation to the Copyright Act 1968.
During 2009 the Monash University ARROW Repository will be loaded with records from the annual research publication data collection managed by the university Research Office and records identified by the Excellence in Research for Australia initiative. These metadata-only records will reduce the percentage of full text content back to 15-20% of the total, which is about the current norm in Australian research repositories.

Monash University staff can add their publications directly, provided they own the copyright in the publication and have the approval of their co-authors to do this. A program to encourage self-submission of pre-prints for the metadata-only records is being prepared, centred on the development of a suitable web form and outreach to the research community on the benefits of providing an open access version of their publications. In addition to this initiative, Library staff will add publisher digital object identifiers (DOI) to the records of applicable publications and link to the catalogue for any book records. While an open access version of the work stored in the repository is preferable, providing these links avoids users searching repositories for actual material and coming to ‘dead ends’.

4.1.1. Theses

The Monash University theses collection is one of the key research collections representing the research activity of the University. Together with other universities in Australia and New Zealand, Monash University is a member of the Australasian Digital Theses (ADT) Service http://adt.caul.edu.au/ and is using its institutional repository to store theses and supply metadata records to the central ADT service using an OAI-PMH provider service. There are currently 683 PhD theses in the Monash University ARROW Repository.

Monash University has mandated the electronic submission of doctoral theses and part of the Monash theses collection dating back to 2000 has been scanned and loaded into the repository. Monash will be adding to the growing number of records on ADT for digital theses, which surpassed 26,000 in September 2008. 24,000 of these have been contributed by members and 2,000 came from the national 50,000 theses records on ADT.

By adding theses to the Monash University ARROW Repository, the University Library hopes to increase the utilisation of the valuable information contained in Monash University theses.

4.2. Research data

4.2.1. The role of the repository

Research data management is the storage and curation of data generated by research. Research data is valuable for a number of reasons. Most importantly, it has value to researchers for the duration of their research and after the completion of the research. It may also have residual value to those researchers after results have been published, as well as having value for other researchers or the wider community. Because of this, and given the investment Monash University and the Australian research community have made in the research, this data needs to be managed in ways that meet statutory and funding bodies’ requirements for present and future research needs (Clarke et al 2009).

In a world increasingly dominated by eResearch, it is no longer satisfactory to just publish research findings and conclusions. Repositories provide a platform to make the data generated by research available as well, allowing other researchers to interrogate the research and thus validating the results. This process also provides an opportunity for further research.

The Monash University ARROW Repository is just one part of the technical infrastructure being developed to support better research data management at Monash University. Other parts of this infrastructure include the Large Research Data Store (LaRDS), collaboration and wiki environments such as Sakai, Confluence and Plone, and tools for managing data assets and metadata (for example, Mediaflux) that can be used by researchers to manage data that is not able to be deposited in the ARROW repository, either because it is not in a final state for publishing orarchiving or because access needs to be restricted to meet legal, ethical or commercial requirements. Through partnership with the Monash e-Research Centre, the Library is involved in all aspects of research data management and ensures that the Monash University ARROW Repository is well-integrated with other parts of the overall technical infrastructure, together with the emerging policies and practices surrounding these.

4.2.2. DARE Project

Monash University is also using its ARROW Repository as a key plank to support the DARE Project. The objective of this initiative is to equip Monash University liaison librarians with the skills and
knowledge to provide advice to researchers on identifying their research management needs and establishing a research data management plan for each significant research activity in the university. The primary purpose of this activity is to ensure that the research data generated by Monash University’s many research programs is being actively managed for long term preservation and for better access by peers and the broader community, but one of the secondary benefits will be to identify many new research data resources collections housed in the departments and schools and move towards adding these collections to the institutional repository.

4.2.3. Case studies

This section describes data management case studies as examples of the role of the Monash University ARROW Repository in data management.

Large datasets - crystallography

There is a pressing need for the archiving and curation of raw X-ray diffraction data in the discipline of protein crystallography. Raw data is critical for validation, methods development and improvement of archived structure, but the relatively large size of these data sets has presented challenges for storage in a single worldwide repository, such as the Protein Data Bank archive.

Crystallography ingest tools were developed by staff from the Library and the Faculty of Medicine. An ingest tool was developed to allow academic staff to prepare the datasets for publication themselves, not requiring technical staff from the Library to review and approve for publication (Androulakis et al 2008). By depositing data in the ARROW repository, crystallography researchers were able to include a permanent citation to the dataset in a peer-reviewed publication (Buckle et al 2007), ensuring that the results could be validated and built upon by others in the crystallography community.

Institutional repositories are relatively stable and adequately funded, ensuring persistence, and are a valuable step in the creation of federated discipline-specific search services. As well as depositing data in the Monash University ARROW Repository, Monash crystallographers have been involved in the development of the federated TARDIS service, which is discussed below.

Sound

Collections of music have also been added into the Monash University ARROW Repository to compliment the collection of research data. There are currently 174 sound recordings from the Kartomi Collection of Traditional Musical Arts in Sumatra and 8 records from the Australian Archive of Jewish Music Collection. Providing access to these collections through the ARROW repository helps the archive achieve its aims of providing “a focus for research and performance of Jewish music in Australia, especially folk and cantorial music and encompassing contemporary Australian Jewish music” (Monash University 2009a) and strengthening local and global links between all those interested in Jewish music.

4.3. Research resources

While the Repository is principally focused on storing and disseminating Monash University’s research outputs, publications and experimental data created by Monash academics, it is also used to support digital collections of unique research source material that have been collected by Monash University over the years. This material is sourced from the schools and faculties and where copyright and the amenability to digitisation permit, the resources are stored in the repository so that the visibility of and the access to the content is greatly improved. These collections of resources support active research by Monash academics, their peers in the wider research community and the general public.

4.3.1. Images

Monash University has a campus in Gippsland which is 150km to the east of Melbourne. The Centre for Gippsland Studies Picture Collection contains more than 3000 photographs dating from the 1860s through to the late 20th century, of which more than 1000 have been digitised and included in the Monash University ARROW Repository. The pictures record some of the history and heritage of Gippsland region of the State of Victoria and are now made more accessible for use by scholars, local historians and the public.

Library staff at the campus worked with the academic responsible for the Monash Centre for Gippsland Studies to digitise the historical photograph collection held by the Centre, which “... document how the landscape has changed through land settlement, farming, industry, mining, floods, droughts and fires” (Monash University 2006).
4.3.2. Newspapers

The digitisation of the Monash University student newspaper, Lot’s Wife, represents an initiative that will bring into the public domain material that would not otherwise be available. This material will further promote Monash University history for research purposes. Lot’s Wife is the student newspaper of Monash University’s Clayton campus, and one of the most renowned and celebrated student newspapers in Australia. It is produced by students, for students and operates as part of the Monash Student Association. Historically the paper reflects Monash University’s radical past and has a reputation for continually providing students with interesting and controversial journalism.

4.3.3. Patents

Scanned copies of patents awarded to Monash University in Australia and internationally are stored in the repository. This collection provides a centrally managed list for Monash University staff to search and allows the dissemination protocols of the repository software to facilitate discovery of the patents through other access portals and search tools, beyond the specialist patent search databases used by the legal community.

4.4. Research publishing services

Monash University operates an electronic press, the ePress (http://www.ePress.monash.edu.au) which provides typical university press services for the online publication of journals connected with the University. The service also provides a print on demand service for a range of University monograph publications. The business model for the ePress is essentially a cost recovery subscription model.

Not all publication proposals meet the ePress’s criteria and some of these journals are referred to the repository unit. By agreement with the ePress, the repository unit offers a free of charge service for storage, online publication and open access dissemination to journal publications that fit the following criteria:

- The publication is connected to Monash University in some way. Usually a Monash University academic is on the editorial board or the leadership for establishing the publication, which is provided by a Monash University school or department.
- The publication is free of charge to the public.
- All the traditional activities of the publisher such as editing and formatting text, managing peer review and copyright compliance are undertaken by the editorial board.
- The repository receives only publication ready material.

Publication directly into the repository offers several advantages; the publication process is managed using desktop tools, takes minimal time (particularly given that the time of academics and research assistants is often donated for this task) and does not require the significant costs of additional webservers and niche publication software. The Australian Government has also changed its requirements for accrediting research income to journal publications; previously only commercially published journals were eligible, but now the criterion is that the publication is peer reviewed. This is a great encouragement to open access academic publication in Australia.

To date the repository is hosting two active open access journals that are published semi-regularly. A total of 100 articles have been published, with over a dozen articles added per year. The Repository Unit does not actively seek new journals, but offers this service to potential candidates when they are identified by the ePress or by data management activities as not suitable for more traditional publication avenues. Such content is delivered to the Repository Unit by email or CD and handled by repository staff like any other article record.

4.5. Format and discipline-specific dissemination

Like many other repositories worldwide, the Monash University ARROW Repository provides metadata records using the Open Access Initiative – Protocols for Metadata Harvesting (OAI-PMH) to search services with a focus on scholarly information, such as Google Scholar (http://scholar.google.com.au), OAIster (http://www.oaister.org/) and the National Library of Australia’s ARROW Discovery Service (http://search.arrow.edu.au/).

The inclusion of research collections in the repository has also enabled format-specific aggregation of Monash University content by search services such as Picture Australia (http://www.pictureaustralia.org/), which harvests the Gippsland Studies Picture Collection, and Music Australia (http://www.musicaustralia.org/) which harvests the Australian Archive of Jewish Music. Both
Picture Australia and Music Australia are highly used, not just by researchers but by the general public.

Monash University is also extending its research impact by providing metadata records to discipline-specific search services such as the following:

4.5.1. Business and economics working papers – Economist Online

Monash University is the only Australian member of the Nereus Network which is integrating access to the economics resources of key libraries, academic publications and other online resources in Europe and beyond. The Monash University Faculty of Business and Economics working papers are being harvested into Economist Online (http://www.nereus4economics.info/), which is a portal for economics. There are 1,288 working papers in the Monash University ARROW Repository being harvested by Nereus.

4.5.2. Protein crystallography research data - Tardis

As well as ingesting crystallography datasets into the ARROW Repository, Monash University staff have led the development of the Australian Repositories for Diffraction Images (TARDIS). TARDIS provides a web-based, federated, discipline-specific search and discovery service for datasets of X-ray diffraction images from the Australian protein crystallography community. A growing group of partner organisations from Australia and the UK are archiving raw data at an institutional level – usually using their local institutional repository – and providing metadata to TARDIS using standard harvesting protocols.

5. Conclusions and future directions

A number of national initiatives and strategic government papers have identified issues that have and will continue to have an impact upon the role of libraries and librarians. The Australian Government has been supportive of the move to open access for publicly funded research and has recognised the need for open access through funding, policy and compliance codes.

Monash University has been an early adopter and lead institution in a number of national co-operative projects this area. At the same time Monash University Library has concentrated on efforts to provide greater access to the University’s research output and material that would not otherwise be available through conventional academic publication.

In terms of learnings and future directions, there is a strong case for further inroads to be made in the addition of research data and datasets into institutional repositories. There is also a developing role for subject liaison librarians as outlined by the DARE Project. Experience and knowledge gained over the next few years will determine the policies, practices, skills, funding and infrastructure necessary to establish good data management practice, both digital and non-digital, in this world of data deluge.

Acknowledgments

The ARROW, DART and ARCHER projects were funded by the Australian Commonwealth Department of Education, Science and Training. The funding was provided through the Systemic Infrastructure Initiative as part of the Commonwealth Government’s strategy Backing Australia’s Ability - An Innovation Action Plan for the Future.

ANDS is funded by the Australian Commonwealth Department of Innovation, Industry, Science and Research (DIISR). The funding has been provided through the National Collaborative Research Infrastructure Strategy (NCRIS) as part of the Platforms for Collaboration Investment Plan.

References


