

TO DARE OR NOT TO DARE

Suzanne Clarke, Monash University Australia, Suzanne.Clarke@lib.monash.edu.au
Wilna Macmillan, Monash University Australia, Wilna.Macmillan@lib.monash.edu.au
Samantha Searle, Monash University Australia,
Samantha.Searle@lib.monash.edu.au
Cathrine Harboe-Ree, Monash University Australia, Cathrine.Harboe-Ree@lib.monash.edu.au
Neil Clarke, Monash University Australia, Neil.Clarke@its.monash.edu.au

ABSTRACT

Monash University has a serious commitment to research data management. This is evidenced by its leadership role of the national Systemic Information Infrastructure projects ARROW, DART and ARCHER, as well as the Monash University Information Management Strategy and specific data management activities. In recognition of its commitment to data management, Monash University was selected to be the lead agent in the Australian National Data Service (ANDS). This is a national project which is working to develop good data management practice, collaboration, infrastructure and services.

This paper will provide background on the strategic environment for data management. It will then provide an overview of the implementation of the data management initiatives at Monash University, which are a collaborative endeavour between the Library, eResearch Centre, Research Office, Information Technology Services and Records and Archives Services.

The Data Management Policy, Procedure and Guidelines will be outlined and various strategies revealed towards the implementation of these. This will include the DARE Project. DARE is not an acronym, but it encompasses the keywords most closely associated with this project: data management, ARROW, archives, research, repositories, enablement, ePress. It is an initiative to increase outreach into the faculties by extending the role of contact librarians to include data management advice, building on their specialist knowledge of the information environments of their disciplines and existing relationships with researchers. A 'triage' model is being developed to enable researchers to manage research data using a range of expertise and infrastructure for current and future needs as the university moves further into the digital age and data management.

A critical component of the Monash University e-research strategy is LaRDS, the university's Large Research Data Store, which provides hundreds of terabytes (TB) of capacity for storage of Monash University research data.

Conclusions will be made about the future of data management and the learnings from the journey outlined above.

INTRODUCTION

Monash University has a serious commitment to research data management. This is evidenced by its leadership role of the national Systemic Information Infrastructure projects ARROW, DART and ARCHER, 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 also the lead agent in the Australian National Data Service (ANDS). This is a national project which is working to develop good data management practice, collaboration, infrastructure and services. From this position of strength and collaboration, both across the University and amongst universities, the role of the contact librarian is being rethought through an innovative approach, called the DARE Project, to increase outreach into the faculties by extending their role.

Monash University is Australia's largest and most internationalised university. In addition to its six Australian campuses, it has campuses in Malaysia and South Africa as well as centres in London and Prato (Italy). Monash University has almost 55,000 students and 7,000 equivalent full time staff. Its current strategic direction is focussing on achieving excellence through a cross-disciplinary, multi-campus, international approach. Consistent with this the university has invested strongly in information technology to support research, teaching and administration.

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 the present and future research needs.

Monash University has established the position of Data Management Coordinator. The position has been newly created to be a primary resource on research data management, policy and planning at Monash University. The Data Management Coordinator works with researchers, e-Research staff and others to support access to, storage, and management of research data. The Data Management Coordinator liaises with a wide range of stakeholders, including the Monash e-Research Centre, ARROW (<http://www.arrow.edu.au/>), ARCHER (<http://archer.edu.au/>) and ANDS Projects (<http://ands.org.au/>) and the University's Large Research Data Store (LaRDS: <http://www.monash.edu.au/eresearch/services/lards/>) to facilitate the identification, storage and management of research data.

Monash University Library has also established the DARE Project to increase outreach into the faculties by extending the role of contact librarians to include data management advice, building on their specialist knowledge of the information environments of their disciplines and existing relationships with researchers. A 'triage' model is being developed to enable researchers to manage research data using a range of expertise and infrastructure for current and future needs as the university moves further into the digital age and data management.

Underpinning the data management efforts are two critical central facilities – the Monash University ARROW Repository and the Monash University Large Research Data Store – which together provide storage for material to be preserved and/or made more widely available.

STRATEGIC DRIVERS

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.

NCRIS (the National Collaborative Research Infrastructure Strategy) and the National e-Research Coordinating Committee have identified data management 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 provide researchers with major research facilities, supporting infrastructure and networks necessary for world-class research.

The *Australian Code for the Responsible Conduct of Research (ACRCR)* was released in 2008 and is to guide institutions and researchers in responsible research practices. The Code (Australia. NHMRC, 2008) replaces the *Joint NHMRC/AVCC Statement and Guidelines on Research Practice* (1997). Amongst other things the code provides advice on how to manage 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”.

The *Strategic Roadmap for Australian Research Infrastructure* outlines where strategic infrastructure investments should be made over the next five to ten years. The inclusion of data management has been emphasised.

DATA MANAGEMENT AT MONASH UNIVERSITY

Monash University decided to break research data management activities into a number of different but related ‘themes’.

Policy and planning

The Data Management Subcommittee of the e-Research Steering Committee oversees strategic planning and provides overall direction. This committee agreed that establishing an advisory group with faculty representation is also required to ensure good governance across the University. A primary task of the Data Management Subcommittee to date has been policy development. At Monash University, policies communicate a commitment to broad principles while related procedures should provide guidance as to how staff and students can enact these principles through their daily work and study.

The Monash University Research Data Management Policy summarises the drivers for improved research data management and the responsibilities of different parties. It is designed to be read quickly by staff at many levels within the organisation

(including senior management) and reflects the University's commitment to complying with the ACRCR.

The accompanying Research Data Management Procedures provide procedural detail and are aimed at researchers. While the Policy articulates broad enduring principles, the Procedures will be subject to regular review as practice and experience informs what works and does not work. At this early stage, it is not possible to define all the processes and desired outcomes. In many cases, the Procedures will simply direct researchers to consult with an appropriate specialist in the Library or another unit. This ensures that researchers are not left at a 'dead end' and consistently raises the profile of available support services.

Communications, advocacy and outreach

Advocating good research data management, systematically communicating its benefits and coordinating outreach activities is not easy. Initial stakeholder analysis identified more than thirty internal and external groups with an interest in Monash University's research data management activities. It is important to work out priorities amongst these stakeholder groups and determine appropriate channels and key messages: communication needs to be finely tuned for each audience, as research data management is not a 'one size fits all' topic.

Some stakeholder groups are very large and complex: the most important stakeholder group, Monash researchers, consists of more than 3,000 academic staff and 4,000 graduate students in countless disciplinary sub-groupings that are spread across ten faculties and multiple campuses! It is planned to reach researchers through a research data management website, an online research data management plan and the use of data interviews to both collect information from researchers and promote the available services. A toolkit consisting of a series of 'quick start' leaflets on various topics is being produced, as evidence from usability studies of help publications indicates that information in small, targeted, easily-digested portions (that can be customised and re-packaged as needed) is more likely to be used by researchers than a single comprehensive document.

Data management in practice

As well as providing advice and guidance, Monash staff involved in data management work directly with early adopters in the research community to improve data management practices and increase uptake of services like the Large Research Data Store (LaRDS), the Monash University ARROW Repository and Records and Archives Services.

The greatest success story to date has been with protein crystallographers. The research team archived x-ray diffraction datasets in the ARROW repository and developed a disciplinary portal TARDIS (<http://www.tardis.edu.au/>) that harvests metadata about the datasets and aggregates it with information from other organisations. This project has increased the impact of the crystallography data originating from Monash University, but more importantly it provides a model for federated data repositories that is being considered by the global crystallography community and may be applicable to other disciplines.

Positive changes do not necessarily require a project as large as the crystallography example. Simply encouraging researchers to move research data from stand-alone PC hard drives and insecure office spaces to centralised secure storage greatly reduces the risk that data will be lost, corrupted or mis-used. Advising researchers before a project starts about ownership and copyright issues and the need to consider future uses of data reduces the likelihood of time-consuming retrospective approval processes. Supporting decision-making on file formats and file-naming protocols can have an impact both during a research project and in the longer term.

As well as working with individuals and small groups, potential changes to structures and business processes with broader impact are being identified. An example of this would be working with the Research Office to incorporate data management planning into existing research administration processes such as research grant applications, human ethics approvals and PhD candidacy confirmation.

Skills and expertise

A large part of Monash University's efforts to improve research data management will be directed towards learning opportunities and incentives for staff and graduate students to improve their research data management skills and expertise.

For the Library this means working with existing providers of research training, such as the Monash Research Graduate School and the Professional Development Unit within the Human Resources Division, to ensure that skills gaps are addressed through events (such as inductions, seminars, workshops and demonstrations) and training resources (such as documentation and podcasts). It also means ensuring that the Library's own staff have the opportunity to learn about research data management. The DARE project (described further below) offers contact librarians the chance to hear presentations from invited guest speakers and to participate in a 'buddy' scheme that supports on-the-job learning.

Leadership and collaboration

As the lead institution in ANDS, Monash University is taking a key role nationally and will work collaboratively with other research organisations in Australia and overseas. Monash University will share its experiences by actively contributing to ANDS forums and other national and regional communities of practice, such as CAIRSS (the CAUL Australian Institutional Repositories Support Service). Staff who are contributing to research data management are also encouraged to network with colleagues from other institutions, particularly in Melbourne.

DARE

The DARE project extends the professional role of contact librarians to include raising awareness of research data management with faculties, connecting researchers with the expertise, tools and services available for good data management practice and providing input to policies and processes. It is a 'triage' model, where the librarian identifies with the researcher(s) the best contacts and tools to manage their data. The contacts may include Records and Archives, the

eResearch Office, the Data Management Co-ordinator, ARROW, ePress, the Copyright Officer and others.

The project builds on contact librarians' information expertise in their subject areas and their positive relationships with researchers developed through research information services, collection building and information literacy programs. Although the project is called DARE, with its implication of daring to take on new roles, it is also sometimes, perhaps more realistically, called the 'little steps' project, which indicates the gradual way skills will be developed and objectives reached.

While the Library played a lead role in the establishment stage of the institutional repository at Monash University, contact librarians were not involved and so informing this group regarding developments in data management within Monash University, nationally and internationally was identified as a first 'little step'. Participation in the project is voluntary and all contact librarians and other senior staff members are invited to participate.

In the same way that Monash University worked with CSIRO (the Commonwealth Scientific and Industrial Research Organisation) and the Australian National University in the establishment phase of ANDS, CSIRO librarians were invited and are participating in DARE. This is mutually beneficial as the approach of the two institutions to research data management is different, but many collaborative research projects are shared. This is a 'little step' towards contact librarians being involved across institutions.

A toolkit of information has been developed to guide contact librarians to key documents that are shared within the Monash e-Research Sakai VRE collaboration environment, where the DARE project is organised. The skills and expertise of the DARE librarians have been enhanced through a series of monthly discussions including: an overview on data management, an introduction to the draft Monash University data management policy and plan and introductions to ARROW, the ePress and copyright. Contact librarians who chose to continue in the program, with the encouragement of a 'buddy' system, began holding conversations with researchers and testing the data management plan in real projects, learning by taking another 'step'. Ongoing monthly discussions include updates on developments in data management, opportunities to meet and hear from others working in different data management roles at Monash University and elsewhere, and opportunities to share experiences and develop the 'triage' model.

A brief discussion guide, developed for Monash University from a model used at Purdue University, guides conversations with researchers through the elements of the research data management plan. Components include:

- A description of the project
- Originators and owners of the data
- Types of data to be collected (formats and origin, digital and non-digital)
- Metadata (schema/standards)
- Format/s of and software used in creation and use of the data
- Volume of data to be managed
- Storage, preservation and archiving of data
- Access policies and provisions

- Retention of research data and records
- Future use/potential of the data
- Existing outputs.

The initial group of 10-12 volunteers held discussions across a broad range of disciplines, faculties and types of data. While the librarians' intentions included exploring the variations in need, these first interviews revealed a broad spread of requirements in types and volumes of data, storage needs, access requirements, formats and more. Most discussions were initiated by the librarians, although some arose from researchers' questions in their ongoing interaction with the librarians. Questions, queries, clarifications, a need for more knowledge and the realisation that this is new territory and many of the questions do not yet have answers, generates interest for both the researcher and the contact librarian.

Initial concern that raising awareness amongst research staff would be a difficult task quite quickly turned to concern that the interest and expectations generated could not be managed within reasonable timeframes. All of the researchers initially approached were interested in exploring data management for their research, even if not immediately.

Common feedback includes:

- processes need to be simple, not too time-consuming
- costs are an issue
- recognition that there are risks in not managing research data
- lack of knowledge of what support is available from the university
- access, ethics, intellectual property, copyright issues
- format complexities
- storage issues, particularly for non digital data.

Contact librarians find data management an interesting extension of their roles, enjoy talking with researchers about their projects, and quickly overcame any initial lack of confidence to take the 'little steps' without first being experts.

Many questions remain at the time of writing: how will the triage model work as data management becomes entrenched in research practice?; how will communication amongst all the players be established?; having made the introductions, will contact librarians be able to have a longer engagement with data management?

The ARROW repository

The Monash University ARROW Repository <http://arrow.monash.edu.au> provides enormous benefits for the University's research community by:

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

Large Research Data Storage (LaRDS)

One of the most common and basic needs with regards to research data management is for secure storage. Many data management problems are associated with storage on local disks (“the C:\ drive”) and portable media such as CDs and DVDs. When choosing where to store their data, researchers may make decisions based on cost and ease of use rather than on the likelihood of technical obsolescence or technology failures, environmental conditions, data security, disaster recovery and the availability of associated technical support personnel and services. Monash University has responded to these issues by offering its researchers affordable, easy-to-use networked storage that is provided by professional Information Technology staff as a shared central service.

The Monash Large Research Data Store (LaRDS) is Monash University’s central petascale research data store. LaRDS provides hundreds of terabytes (TB) of capacity and can be used to store all types of Monash research data, including raw experimental data; data downloaded directly from scientific instruments; backup of research data held on the c: drive; databases; simulations; collaborative works; cultural archives; audio, video and multimedia digital content; long-term preservation of archival material; and published research results via the ARROW repository.

LaRDS is available for use by all Monash researchers, including HDR students. It can be accessed via a number of different connectors suitable for different applications and modes of usage and can also be accessed remotely (from other research institutions, home or worldwide) via the high performance computing facility, the my.monash portal, the NetStorage web service, the Monash e-Research Sakai VRE collaboration environment and a variety of other purpose-built systems and web applications.

Data held in LaRDS is reliably backed up and secure. Access to LaRDS data can be limited to individuals, workgroups or more broadly as appropriate..

CONCLUSIONS, LEARNINGS AND FUTURE DIRECTIONS

Government policy has recognised the need for data management through funding, policy and compliance codes. Monash University has been an early adopter and lead institution in this area.

Data management is rapidly gaining momentum in academic and research institutions and this opens up opportunities for librarians to expand their roles, skills and expertise by working with researchers, information technology specialists and records managers and archivists to manage the rapidly increasing amounts of data generated by research.

There is significant interest in this area from all the players, who are keen to expand their knowledge and roles. Researchers have welcomed discussions and recognise the importance of data management even when unsure about pathways forward.

Strategies and knowledge gained over the next few years will determine 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 have been funded by the Australian Commonwealth Department of Education, Science and Training. The funding has been provided through the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovation Action Plan for the Future.

ANDS is funded by the Australian Commonwealth Government's [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.

Peter Mathews, Library Planning Executive, Office of the University Librarian, Monash University Library, developed the DARE Toolkit and Discussion Guide.

REFERENCES

Australia. Department of Education, Science and Technology (2007). *Platforms for collaboration*. Retrieved October 10, 2008, from http://ncris.innovation.gov.au/capabilities/collaborative_investment_plan_platforms.htm

Australia. Department. of Innovation, Industry, Science and Research (2008, August). *Strategic Roadmap for Australian Research Infrastructure*. Retrieved October 10, 2008, from <http://www.innovation.gov.au/ScienceAndResearch/Documents/Strategic%20Roadmap%20Aug%202008.pdf>

Australia. National Health and Medical Research Council, Australian Research council (2008). *Australian Code for the Responsible Conduct of Research*. Retrieved October 10, 2008, from <http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>

Cutler, T. (2008). *Venturous Australia: Building Strength in Innovation*. Retrieved October 10, 2008. from <http://www.innovation.gov.au/innovationreview/Documents/NIS-review-web.pdf>

Harboe-Ree, C., Treloar, A. & Groenewegan, D. (2007)., Data management and the curation continuum: how the Monash University experience is informing repository relationships. *D-Lib Magazine*, 13 (9/10). Retrieved October 10, 2008, from <http://www.dlib.org/dlib/september07/treloar/09treloar.html>

HLB Mann Consulting. (2007). *Evaluation of Library Help Publications* [consultation report].

Lougee, W. P. (2002). *Diffuse Libraries: Emergent Roles for the Research Library in the Digital Age*. Council on Library and Information Resources. Retrieved October 10, 2008, from <http://www.umanitoba.ca/libraries/engineering/mla/difflibs.pdf>

Lynch, C. (2008). Big data: How do your data grow? *Nature*, 455, 28-29.

Macdonald, S. & and Martinez-Urbe, L. (2008). *Data Librarianship: a gap in the market*. Retrieved October 10, 2008, from <http://www.cilip.org.uk>

Monash University, (2005). *Information Management Strategy*. Retrieved October 10, 2008, from <http://www.monash.edu.au/staff/information-management/>.

Nature Magazine (2008). The Data Deluge. *Nature*, 455,1-136.

Palmer, C., Tefteau, L. & Newton, M.P. (2008). *Identifying Factors of Success in CIC Institutional Repository Development, Final Report*, Retrieved October 10, 2008, from <https://www.ideals.uiuc.edu/handle/2142/8981>

Swan, A. & Brown, S. (2008). *Skills, Role and Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs*. Retrieved October 10, 2008, from <http://www.jisc.ac.uk/publications/publications/dataskillscareersfinalreport.aspx>

Witt, M. & Carlson, J.R. (2008). *Conducting a data interview*. Retrieved October 10, 2008, from http://docs.lib.purdue.edu/lib_research/81