Introduction

Monash is a large, research intensive Australian University, with 6 campuses in Victoria and 2 overseas. It has over 120 research centres and institutes, across a wide variety of disciplines. Monash has a history of strong leadership in e-Research and scholarly communications, particularly in research data management (RDM), and has been instrumental in pioneering initiatives such as the Australian National Data Service (ANDS).

This case study looks at the infrastructure and services in place at Monash University to improve research data management. The main section is broken down into seven themes as follows:

1. Data management strategy
2. Data management policy
3. Guidance and training
4. Research data storage and archiving
5. RDM platforms
6. Metadata
7. Data Management Planning

We describe the approach taken by Monash under each theme and discuss how different elements have evolved and why, to provide examples that other institutions may wish to adopt.

Roles and responsibilities

Research data management requires coordination and collaboration across multiple stakeholders. Monash University has succeeded in achieving broad buy-in across the different services, with high-level commitment from the Deputy Vice Chancellor for Research (who is also Monash’s Provost), and excellent researcher engagement. A Research Data Management Subcommittee was first convened in 2006 to lead and co-ordinate this work. The subcommittee continues to develop and is now reporting to the Research Committee. There is a recognition that relationships with researchers are paramount; an emphasis on researcher engagement and building trust is apparent throughout the various RDM initiatives.

The Library, Monash e-Research Centre (MeRC) and eSolutions (the central Information Technology division), working with researchers, are the main players in developing and implementing Monash University’s RDM strategy. The Library has good relationships with researchers and research managers and is a trusted partner in research. Library staff have many skills that can be applied to improving RDM practices. The Library has also been the locus for activities relating to scholarly communication. MeRC is similarly well-placed to improve RDM practices for the University. They work in close partnership with researchers to develop disciplinary RDM platforms that address their issues and needs. The MeRC team adopts an agile, scrum-based approach to development that is researcher-led. eSolutions play a critical role too. They provide the generic storage infrastructure and support the disciplinary RDM platforms post-development, once they have become suitably robust.

This user-orientated, collaborative approach has seen Monash advance thinking in the RDM field.
Research Data Management at Monash

Monash University has been addressing research data management since 2006. In this time an overarching vision has emerged. The following sections will describe activity in different areas and attempt to place these elements in the overall context, as seen in Figure 1.

1. Data management strategy

RDM initiatives at Monash grew out of an information management strategy, which was released in 2006. Concerns about information management issues came to the fore at the 2002 Monash University Information Technology Strategic Planning Retreat. This led to the establishment of a broad steering committee to address these issues. The ensuing strategy covered the management of all University information across three areas: administration & support; learning & teaching; and research & research management. In terms of research support this entailed developing research collaboration spaces and improving the management of datasets, amongst other things.

A Research Data Management Strategy and Strategic Plan has been released for the period 2012-2015. This puts forward five data management themes, aligning each with other University strategies to show how research data management contributes to the University’s research, education and professional objectives. Future RDM initiatives are planned under each of the five themes with concrete goals, key measures and responsible partners being clearly assigned. Notably, the strategy is Creative Commons licensed to promote reuse and has already been emulated by the University of Bath.

2. Data management policy

RDM policy and procedures were developed over several years of engagement with researchers. Monash’s current Research Data Management Policy was approved in 2010. It responds to the Australian Code for the Responsible Conduct of Research by articulating the University’s commitment to research data management as a shared responsibility between researchers, academic units and central administrative units.

Figure 1: Institutional Infrastructure for Research Data Management

---

Early iterations of the policy and procedures experimented with organising the information in different ways, for example by responsibility. In line with all Monash University policies there is a 3-tier approach:

1. **Policy** - a high-level set of principles, approved by the Academic Board and Council that is expected to remain stable

2. **Procedures** - guides identifying roles and responsibilities of the different groups that can be updated and agreed through the Monash University Research Committee

3. **Guidelines** - flexible, regularly updated advice delivered via the University website

A number of challenges were encountered while developing the policy and procedures. The process involved consultation to: suit the legislative contexts of each campus in Australia, Malaysia and South Africa; integrate with other University research related policies; and ensure applicability to a broad range of research practices. Agreeing on definitions was particularly controversial. They decided to use the process as a communication opportunity to raise awareness of research data management, develop effective partnerships with key stakeholders and achieve a shared understanding on terminology and recommended approaches. This also lead to the practice of communicating the benefits to researchers rather than framing data management in terms of risks and compliance.

Whether policies should pre-date infrastructure is a moot point. Without the infrastructure a policy can be complex to implement, but without the policy it can be hard to leverage investment to develop the infrastructure. Monash University chose to address both simultaneously, continuing to build the infrastructure and capability needed while developing the policy framework to demonstrate the institution’s commitment to improving RDM.

### 3. Guidance and training

The development of the research data management guidance webpages at Monash started in parallel with the policy and procedures. The webpages have been available since May 2009 and continue to be updated to provide detailed and practical advice in everyday language. Topics include data management planning, ownership, copyright & intellectual property, ethical requirements and storage & backup. The website has been visited over 18,000 times and traffic from Monash to ensure a range of options are available to meet different audiences and needs. Two-hour data planning seminars aimed at new postgraduate research students have been run, along with induction and workshop sessions (particularly within faculties and schools), e-Research seminars, workshops and network breakfasts. There is an emphasis on tailoring skills development; individual consultations are available from Library and MeRC staff and workshops can be organised on request.

Effort has also been made to ensure that data management skills and knowledge bridge the University’s research and teaching environments. A collaboration between the data management co-ordinator, learning skills advisers and contact librarians is applying the toolkits and expertise from the information research and learning skills portfolio to the emerging area of RDM. In the future there will be an increased emphasis on embedding training within faculty-based coursework and professional development programs and in working with academic staff to develop more integrated approaches. Skills and knowledge is one of five themes in the RDM strategy and a number of goals have been set to develop and integrate RDM skills development for both researchers and professional staff.

### 4. Research data storage and archiving

Established in 2006, the Large Research Data Store (LaRDS) is the preferred storage environment for research data at Monash University. LaRDS is a petascale research data store, providing thousands of terabytes of storage capacity. It provides a reliable, secure, long-term common data storage infrastructure. All information is automatically backed up to tape in the University’s two physically diverse data centres.

LaRDS is freely available to all Monash researchers, including postgraduate research students, and can be used to store all types of research data. Access to data is controlled by the user and can be restricted to individuals or workgroups, or made more broadly available as appropriate. LaRDS is made available to a wide spectrum of applications and services, so as to meet the diverse needs of different research groups. Case studies show how it has been used for routine data back-up, to collaborate and share files using Confluence wiki and the Sakai virtual research environment, and to publish data via the Monash Research Repository, ARROW.¹⁰

The Monash University Research Repository contains content representing Monash University’s research activity. It was initially developed as an open access repository.
publications repository but has since been extended to expose research data holdings. The repository provides a place to securely store and centrally manage selected research data, collections, and related publications so they are globally accessible online. It contains: accepted versions of published works like books, book chapters, journal articles and conference papers; non-published manuscripts and grey literature like theses, technical reports, working papers, and conference posters; and research data holdings, including data sets, image collections, audio and video files.

5. RDM platforms

Arguably, what differentiates Monash from other universities is its approach to developing data management capability and platforms during the active phase of research. They have adopted a federated, disciplinary approach to research data management via RDM platforms, as per the lifecycle diagram below. There is a recognition that the broad-based, enterprise solutions which IT services excel in providing are often not best suited to researchers needs. For an RDM platform to be effective, it needs to fit in with a researcher’s practice, their instrumentation, research tools, IT environment and culture. Most of these features vary from discipline to discipline, so it is unrealistic to believe that a singular approach will consistently meet researchers’ needs. As such, MeRC partners with research groups to develop RDM solutions tailored to specific communities.

An important mantra is followed when developing RDM platforms. If a research community already has a solution (or there is an emerging one), they adopt this and, where necessary, adapt it to suit the needs of researchers at Monash. Only as a last resort will an entirely new solution be developed. Developing a new product may be expensive, costly to support, and could split researchers from their community. This approach acknowledges that researchers’ loyalty is often stronger to their discipline than their institution.

Enhancing existing RDM platforms improves sustainability as the solution is owned by a research community rather than the institution. Monash has found that researchers embrace the solution and promote it within their community, encouraging their collaborators to enhance and extend it. In several cases researchers have secured additional funding to maintain and improve their RDM platform.

MeRC adopts agile software development methodologies with at least one researcher as a product owner to guide development. Funders are usually included in the team to ensure the approach remains flexible and can make unanticipated changes of direction to meet users’ needs. The disciplinary RDM platforms are developed by MeRC in collaboration with researchers in what’s informally known as the ‘Chaotic Sand Pit’. When a new platform is ready for to be released in a production environment, it is released in the ‘Healthy-Hot-House’ (a nursery for new research systems). Once the platform is mature, it is then passed on to eSolutions to manage.

MyTardis is just one of the programs developed in this way. Crystallography researchers were at risk of losing data as they had to carry external hard drives to facilities and transport their data home after running their experiments. MyTardis was developed to record the data generated from an experiment, catalogue it, and transfer it back to the home institution. It is now in use at the Australian Synchrotron and many institutions around Australia. MyTardis has also been adapted for Electron Microscopy. Similar systems have been adopted and further developed for other disciplines, such as OMERo for Optical Microscopy. These platforms allow researchers to capture, store, organise, search, share and publish their data. MeRC also provides basic support for Research Data Management by offering researchers: disk mounts on LaRDS; Sakai (a virtual research environment); and Confluence (an enterprise wiki).

1 Diagram from a presentation made by Anthony Beitz at Open Repositories 2012
2 Details on Tardis and MyTardis - http://tardis.edu.au
From its continued success in e-Research and its researchers’ growing demand for it, Monash University has managed to grow a sizeable e-Research team. This places Monash in a good position to offer customised solutions to some research disciplines. However, such an approach may not be feasible in all institutions. Some may prefer to deliver generic institution-wide services that meet the majority of use cases, so that all researchers have some level of provision. Generic services are more achievable in the face of budget cuts and calls for efficiency, especially given the push from research funders to have appropriate infrastructure in place. They may not be the most cost-effective in the long-term though. The Monash approach of tailored research data management is worthy of investigation given the enhanced levels of user satisfaction, buy-in and sustainability that ensue. As Simon Hodson concludes in his blog post on this subject: “there is little point in providing generic solutions if these do not respond sufficiently to researchers’ requirements and are scarcely fit for purpose.” In the long term, a disciplinary approach that involves collaboration across national and institutional boundaries, and most importantly is led by research communities, may be more sustainable.

### 6. Metadata

Australia aspires to maintain a catalogue of research data in an accessible form, as outlined in the Australian Code for the Responsible Conduct of Research. A lot of work has been initiated by ANDS to develop this. Research Data Australia (RDA) is a national discovery service for research data, populated with metadata harvested from Australian institutions. The ANDS ‘Seeding the Commons’ programme has funded many universities to identify research data collections and to catalogue them in an institutional research metadata repository, from which RDA can be seeded. Monash University has decided against developing an institutional research metadata repository at present, as there are concerns that this capability can be provided at the national level, saving the individual institutions from duplicating infrastructure. The various RDM platforms in use at Monash already allow research data to be described, so metadata is exported from these to the appropriate national, international, and community services such as Research Data Australia and MyTardis (for raw protein crystallography research data). These relationships can be seen in the ‘discovery infrastructure’ layer of Figure 1.

### 7. Data Management Planning

Monash University guidelines encourage all researchers to undertake data management planning at the start of each research project. The library developed an initial data plan template in 2007 and trialled this with researchers. They found that researchers needed assistance negotiating statutory requirements and intellectual property issues, so produced a revised template that was used by librarians to conduct data interviews. This validated the findings from 2007 with more researchers across a broader range of disciplines.

The current multiple choice checklist walks researchers through all the aspects of data management covered by the Australian Code for Responsible Conduct of Research and Monash University’s Research Data Management Policy. It reflects a change in approach taken in 2009, outlined in the table below:

<table>
<thead>
<tr>
<th>Previous version/s</th>
<th>New version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan = output</td>
<td>Planning = a process</td>
</tr>
<tr>
<td>Focus on compliance</td>
<td>Focus on self-assessment and discovery</td>
</tr>
<tr>
<td>Form, open questions, free text response</td>
<td>‘Checklist’, more multi-choice + space for free text</td>
</tr>
<tr>
<td>Assumed: researcher knows the answers but they are not written down</td>
<td>Assumes: some topics may be new to the researcher or not well understood</td>
</tr>
<tr>
<td>Stand-alone contacts</td>
<td>Direct links to policies, web resources services,</td>
</tr>
<tr>
<td>Documents existing practice</td>
<td>Promotes best practice</td>
</tr>
<tr>
<td>Duplicated existing sources</td>
<td>Suggests existing sources are attached or referred to</td>
</tr>
</tbody>
</table>

---

14The data planning checklist is available at: http://www.researchdata.monash.edu.au/guidelines/planning.html
15Table from Summary of data management activities at Monash University 2006-2011, [internal document]
Data management plans are not required by Australian funders as they are in the UK, giving institutions greater flexibility on how they implement them. At Monash, plans are encouraged rather than mandated; the Library and MeRC are careful to maintain the trusted relationships they’ve built with researchers and are wary of imposing additional work. Without a mandate from funders, requirements will be kept to a minimum and questions will only be asked if they lead to direct benefits. Further work is forthcoming to investigate and deploy data planning methodologies and tools. Over time, Monash aims to scale up existing data planning activities to target a larger number of researchers and investigate ways to automatically capture information to feed into the planning of infrastructure, training and other programmes.

Conclusion

There are several areas where Monash University has led the field in terms of research data management. Their policies, strategies and guidance were developed far earlier than others and have provided models that institutions around the world have referred to and emulated. The technical infrastructure at Monash is similarly pioneering, with the institution taking part in or leading major national initiatives. The LaRDS research data storage facility, for example, has been in place since 2006 at Monash, whereas many UK institutions are only now starting to offer high-volumes of research data storage as a core provision. What really sets Monash University apart though is its bold approach to developing RDM platforms. Focusing predominantly on providing bespoke platforms in conjunction with capability building, tailored to specific research group needs, is a sea-change from the common approach of providing institution-wide RDM solutions. Understandably, most institutions may find this approach too risky, however it does open up opportunities for cross-institutional, discipline-focused partnerships and shows much more promising signs for sustainability.

There are a number of important lessons to take from the case study. Firstly, the length of time Monash University has been working on research data management. The initial strategy and introduction of RDM storage date back to 2006, showing how long it takes to build momentum. Each area of work has slowly evolved over time, requiring continual effort. It is not sufficient to address something once and assume the job is done. Secondly, there have been a few areas where Monash was unsure of the best approach to adopt, notably in terms of metadata and data management planning. They’ve had the confidence to hold back and trial different options rather than commit to a model they weren’t sure was in researchers’ interests. And thirdly, the greatest successes have occurred because of researcher involvement. Consultation and partnership working has predominated, enabling them to secure the buy-in needed for uptake and sustainability. The research data management challenge is far from solved at Monash University; the strategy for 2012-2015 shows how much work is still to be done. However, the time and resources committed over the past six years have broken new ground and offer a number of useful examples that can be adopted by other institutions.

Acknowledgement

This case study is informed by discussions with staff at Monash University and various papers and presentations that they have shared. We are very grateful to Sam Searle, Anthony Beltz, Wilna Macmillan and Andrew Harrison for their input and corrections.

Follow DCC on Twitter: @digitalcuration and #ukdcc