

The International Journal of Digital Curation

Volume 8, Issue 2 | 2013

Adding Value by Taking a National and Institutional Approach to Research Data: The ANDS Experience

David Groenewegen,
Director, Research Infrastructure, Monash University Library
Monash University

Andrew Treloar,
Director, Technology,
Australian National Data Service

Abstract

The Australian National Data Service (ANDS) has been working to add value to Australia's research data environment since 2009. This paper looks at the changes that have occurred over this time, ANDS' role in those changes and the current state of the Australian research sector at this time, using case studies of selected institutions.





Introduction

In this paper we provide an overview of the state of data management in the Australian research producing sector, especially as it relates to universities and other large institutions. We also look at ANDS' role in the sector and some recent initiatives.

The Australian National Data Service

The Australian National Data Service (ANDS)¹ was established in January 2009. ANDS exists to transform Australia's research data environment by making Australian research data collections more valuable through managing, connecting, enabling discovery and supporting the multiple use of this data. The purpose of this activity is to enable richer research, more accountable research, more efficient use of research data, and improved provision of data to support policy development. The outcome of this activity is that Australia's research data as a whole are steadily becoming a nationally strategic resource.

Substantial infrastructure has already been created and ANDS has a range of national services in place, including a data collections registration service, a data collection description publication service, a data citation service, a researcher identification service, a "see-also" service that enables other discovery tools to use/be used by the ANDS discovery service, a research activity identification service, a research project identification service, and the enhanced Research Data Australia² (a data collections discovery service).

This is combined with a range of coherent institutional research data infrastructures, including tools deployed to automatically capture rich metadata along with the data for a wide range of instruments, operational metadata stores for this metadata, and collection description feeds to ANDS from both research institutions and public sector data holders. Institutions responsible for nationally significant collections have been supported in making those collections connected, visible and available, often through a Research Data Storage Infrastructure³ node; around 90,000 collections are available for discovery through Research Data Australia and discipline-oriented portals are cross connected to Research Data Australia.

The following tools, frameworks and capabilities are in place to further exploit this infrastructure:

- Improved institutional guidance for internal institutional data management,⁴

¹ Australian National Data Service: <http://ands.org.au/>

² Research Data Australia: <http://researchdata.ands.org.au/>

³ Research Data Storage Initiative: <http://rdsi.uq.edu.au/>

⁴ See ANDS Data Management: <http://www.ands.org.au/datamanagement/index.html>

- 
- Improved institutional guidance for responding to national instruments such as the Australian Code for the Responsible Conduct of Research (NHMRC, [2007](#)),
 - Institution-wide research data management planning frameworks or improved research data management at many research institutions,
 - Increased institutional capability for research data management with staff trained in this area,
 - Tools that enable more effective re-use of research data⁵.

ANDS has now received funding to continue a subset of its activities until mid 2014, as well as anticipated funding for service and activity maintenance through to mid 2015.

Moving from Funded Projects to Institutional Engagement

Much of the infrastructure described above (with the exception of the national services) has been funded by ANDS in the form of activity and software development at research producing institutions. This funding was derived from a number of Commonwealth Government programs that are now concluding. As a result, all of the current round of externally-funded programs should have concluded by the end of 2013.

This means that ANDS needs to continue to transition its interaction with research producing institutions from a project-funding model to one focused on provision of expertise and advice (as well as a limited amount of effort). Under this approach, ANDS is working to use the infrastructure described above to build on existing relationships with research institutions to help them to further achieve their research data management ambitions and needs. This engagement can take a number of forms, including formally agreed programs of work, regular placement of ANDS staff at partner locations, ANDS participation in planning, and provision of informal advice and guidance in areas such as software and policy. ANDS sees its role as bringing together and disseminating the best and most current advice, services and tools, and helping to implement them as required, based on the needs of the local partner. How we approach this is influenced by similar initiatives as those offered by the Digital Curation Centre.⁶

A Bidirectional Traffic in Influences

The Australian Higher Education sector does not exist in a vacuum, and trends in funding overseas and the need to participate in international collaborations mean that local institutions need to meet the commitments made by their overseas partners. As more overseas funding bodies require data management plans to obtain grants, the expectations on local institutions are increased. ANDS is also increasing its range of

⁵ See ANDS Project Registry: <https://projects.ands.org.au/>

⁶ See: <http://www.dcc.ac.uk/community/institutional-engagements>

international engagements to ensure that Australian research can be compatible and connected.

Of particular impact is the push towards improved institutional data management practices, driven by funding bodies. The US National Science Foundation (NSF) requirement that grant recipients must provide a data management plan (NSF, [n.d](#)), and the UK Engineering and Physical Sciences Research Council (EPSRC) policy framework on research data (EPSRC, [n.d](#)) have been the subject of considerable interest in ANDS' engagements with local research institutions. In our engagements, the news that these requirements are in place seems to generate increased enthusiasm for the need to improve institutional data management ability. The EPSRC policy framework has had the added benefit of providing new examples of how to do this for local institutions. There is a pleasing symmetry to this, as at least one UK roadmap is based on an Australian one produced in 2011 (Pink, [2012](#)).

Further international collaboration and impact has been seen in the use of data management planning tools. The UK and US have been working together on the development of an online tool for some time, and in response to requests from Australian institutions ANDS has developed a local variant, deployable in the cloud, based on the DMPOnline tool developed by the Digital Curation Centre⁷. Interest in this type of tool seems to be driven partly by ANDS-funded projects, and partly by the recognition that if data management is to be attempted at an institutional scale, then it needs to be automated in a sustainable way. Emailing Word documents around so they can be printed and “put somewhere” is not an option.

The interest in overseas funders requiring data management planning also seems to have had the effect of increasing the interest of local funding bodies in proposing similar requirements. Australian funding councils have thus far not been prepared to mandate data deposit or data management planning, but as it becomes more prevalent with our research partners, it seems that these attitudes are changing (Palmer, [2012](#)).

As well as Australia being influenced by overseas practice, Australia has been influencing some aspects of international activity. Australia (through ANDS), together with the United States of America (through the NSF-funded RDA/US project) and the European Union (through the RDA-Europe (formerly iCORDI) project) have brought about the formation of the Research Data Alliance⁸. The purpose of the Research Data Alliance is to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability. This will be achieved through the development and adoption of infrastructure, policy, practice, standards, and other deliverables. The Research Data Alliance was launched in March 2013 in Gothenburg, and held its second plenary meeting in September 2013 in Washington DC.

Trends in Institutional Data Management in Australia

As a result of the changes within the environment noted above, ANDS has observed a number of trends in the way that data management is approached by Australian

⁷ DMPOnline: <http://www.dcc.ac.uk/dmponline>

⁸ Research Data Alliance: <http://rd-alliance.org/>

research producing institutions, and has been seeking to encourage and support them. These include an increase in the level of eResearch and data support resourcing; a growth in the creation of relevant IT infrastructure, such as storage and support for metadata stores that describe the data; a wider embrace of the need for planning around the creation of research data and the need for wider internal organisation to support this.

When ANDS began there were only a handful of staff dedicated to research data management working in Australian research institutions, who themselves were generally still grappling with the concepts and challenges involved. ANDS' project funding has substantially increased the numbers of staff, as would be expected, and increasingly institutions are using their own resources to fund work in this area. ANDS has identified ten universities (out of the 40 in Australia) who have put an average of A\$300,000 into ongoing annual investment in this area. Selected universities are discussed in the institutional case studies below.

Areas of particular investment include the creation of data management co-ordinator roles, implementation and support of 'metadata stores' that describe the data holdings of the institution, new data storage facilities, training programs in data management principles, creation of eResearch support and software development centres and integration of institutional sources of truth in single systems to better track research. These tools and services are then brought together to provide greater institutional coherence in the broader support for research. This is helping to make more data widely available.

Additionally, as data becomes more widely available, researchers are exploring what you can do with data and are increasingly looking for new services to better use the data and collaborate over it. Many of these services are starting to be made available in the Cloud. The need for these new services is increasing as the sizes of data collections increase, meaning that the old paradigm of moving the data to the user is impracticable.

Examples of Data Sharing

In recent years ANDS has seen some positive results from data sharing that reinforce some of the assumptions made when we began funding work in this area. While these examples, funded in part by ANDS, are still isolated, we believe that they indicate the positive effects of better data management and the sharing of that data.

Parkes Observatory Pulsar Data Archive

Shortly after the release of the Parkes Observatory Pulsar Data Archive⁹, a CSIRO researcher gave a presentation about it to Chinese astronomers while on a visit. They have used the data to look for glitch events in pulsar data sets, and the results of this work, undertaken jointly with CSIRO, are soon to be published. The Archive is also serving as a major resource in an international search for gravitational waves. Therefore, by managing and making the data available, new research has been enabled and new research collaborations have been formed.

⁹ Parkes Pulsar Data: <http://atoa.atnf.csiro.au/parkesdata.jsp>



Repository of Antibiotic Resistance Cassettes

The Repository of Antibiotic Resistance Cassettes¹⁰ is a system at the University of New South Wales that allows microbiologists to browse and explore the gene cassette repository, annotate cassette array sequences using the knowledge base and the Attacca annotation engine, and contribute new cassettes not yet in the database and obtain unique names for them. Within six months of its launch it was instrumental in the discovery of more than 40 new genes, including about ten that may confer antibiotic resistance and are thus clinically important. It now services users in 11 countries, primarily Australia. It is going to remain an important tool in the global fight against antibiotic resistant disease for some years yet.

Life Patterns

Life Patterns¹¹ is a project of the University of Melbourne's Youth Research Centre. Making their longitudinal studies available through the Online Heritage Resource Manager (OHRM) run by the University of Melbourne has created new opportunities, including collaborating with a Swedish postdoctoral fellow to undertake a new analysis of the data and a successful ARC Linkage project with new partners.

Redmap

Observations of marine organisms by non-scientists can be added to traditional scientific sources using Redmap¹², a service hosted by the University of Tasmania. Not only does this service allow 'citizen scientists' and professional researchers to create a verified sightings resource, the data captured has opened up new research collaboration opportunities. In one example, a Mexican graduate student emailed about the possibility of doing similar work in Australia. The researcher had just picked up an unusual observation of a gloomy octopus (*Octopus tetricus*), traditionally found off Sydney and the east coast of Australia but not in Tasmania where it was sighted. The graduate student is now studying this new phenomenon.

AusNC

Linguistics researchers from across the country have come together to create the Australian National Corpus¹³: a discovery service that collates and provides access to assorted examples of Australian English text (published and unpublished), transcriptions, audio and audio-visual materials. The establishment of the AusNC has also seeded collaborative projects, including the development of a virtual Human Science Communication lab and the deployment of language data from the AusNC in educational contexts. New research projects include the Irish influences on Australian English and further developing an interlingua ontology for use across spoken data collections.

¹⁰ Repository of Antibiotic Resistance Cassettes: <http://rac.chi.unsw.edu.au:8080/rac/>

¹¹ Life Patterns: http://web.education.unimelb.edu.au/ycr/life_patterns/

¹² Redmap: <http://www.redmap.org.au/>

¹³ Australian National Corpus: <http://www.ausnc.org.au/>

While these services have largely been driven by the needs of the disciplines or the researchers within them, ANDS has also been watching and encouraging the development of increased institutional services and support for researchers and their requirements.

Institutional Case Studies

The following case studies describe how selected Australian universities are embracing these challenges. This is not an exhaustive study of all the work being done, and is intended to give a flavour of the activity underway. All of the case studies chosen are characterised by substantial ongoing commitment of internal resources to data management (such as dedicated staff, senior buy-in at governance level), a whole of institution view of the issue (rather than trying to solve a bit of the puzzle), the use of substantial IT infrastructure related to data management (such as a well-connected metadata store and large scale storage) and ongoing outreach to researchers at a variety of levels.

Melbourne University

Melbourne University provides comprehensive support for researchers to enable consistent management of research data and materials in line with the institution's policies and guidelines. To encourage the uptake of research data management by researchers within the university, it was decided to create well-resourced central services that would support research data management. The Research Data Management Toolbox¹⁴ has been developed by the library working with Information Technology Services and the research office. The toolbox supports departments, researchers and research administrators in implementing University policy. It includes Data Management Plan templates and links to relevant documentation and contact addresses. The Central Research Data Registry is an institution-wide record of the research data and records stored at the University of Melbourne. The central registry includes a description of the research data and records, the name(s) of associated researchers and projects, the location of the data (digital and analogue), access restrictions, relocation and disposal schedules, and the organisational unit responsible for the ongoing stewardship of the records and data.

Edith Cowan University

West Australia's Edith Cowan University has a policy that the researcher is responsible for the management of research data and materials. The library offers a first point of contact with its email researchonline@ecu.edu.au, which handles research data enquiries and directs them to the appropriate person, and with the resources at their Research Data Management site.¹⁵ The Graduate Research Induction Program (GRIP) offers training in data management principles. There is no central storage dedicated to research data and requests are handled on a case-by-case basis.

¹⁴ Research Data Management Toolbox: http://imas.unimelb.edu.au/research_data_management

¹⁵ Edith Cowan University Research Data Management: http://ecu.au.libguides.com/research_data_management



Griffith University

Griffith University has been working on data management for a considerable number of years and had invested in this prior to the formation of ANDS. ANDS has acted as a catalyst and shaper of the Griffith data management vision. This vision continues to undergo review. The data management agenda has been and continues to be driven primarily by the eResearch Services Team and management in Information Services, as well as increasingly in collaboration with the Office for Research. Information Services is a centralised service which encompasses the IT, eResearch and Library sections, which is a distinct advantage in actioning things. The university has made available a number of advisory and support documents on their eResearch services page¹⁶ and describe and share their research data through the Griffith Research Hub.¹⁷

Monash University

Monash University is committed to improving the way research data in all formats is created, stored, managed and disseminated. They have made substantial resources available on the Research Data Management section of their website¹⁸, and have a publicly available Research Data Management Strategy and Strategic Plan 2012-2015 (Beitz, Dharmawardena and Searle, 2012). Work in this area is a collaboration between the Library, eResearch Centre, eSolutions and the Research Office. Research staff and students at Monash University have access to a wide range of services and tools for managing research data. For example, the university library provides support for researchers who are completing the checklist and can also tailor meetings and workshops for groups of researchers on request. The Large Research Data Store (LaRDS) offers central storage for research data. This storage is permanent, secure, scalable, and backed up daily to two data centres in different physical locations.

University of Newcastle

At the University of Newcastle researchers and departments are responsible for the management of the research data and materials. The institution provides support for the researchers to enable consistent management of research data and materials in line with the institutions policies and guidelines. These were some of Australia's earliest specifically created policies and procedures, and include a Research Data and Materials Management Policy, a Research Data and Materials Management Procedure, a Research Data Storage Facility Procedure and a Responsible Conduct of Research Policy. The library can assist with research data management and planning. This includes advice, consultation, referral and provision of tools and resources.

The research office, library and IT work in partnership to provide advice and support to researchers. The library provides support for the development of research data profiles, including citations and their inclusion in Research Data Australia and NOVA (The University of Newcastle's digital repository), publishing of supplementary research data for journal publications, and the creation of Researcher

¹⁶ Griffith University eResearch: <http://eresearch.griffith.edu.au/>

¹⁷ Griffith Research Hub: <http://research-hub.griffith.edu.au/>

¹⁸ Monash University Research Data Management: <http://www.researchdata.monash.edu.au>

Identifiers with the National Library of Australia service. A Data Management Toolkit¹⁹ is available to provide information and resources to assist with planning, managing, storing, sharing and publishing research data. Faculty librarians have been trained in research data management with the intention of making them the contact point for researchers who are responsible for managing their own data and research materials. The University of Newcastle's Research Data Storage (RDS) facility provides for the secure storage of research data for University of Newcastle researchers.

Conclusion

Research data management has progressed considerably in Australia over the past four years and ANDS has been central to making this happen.

As can be seen in the examples above, there is some pleasing progress at an institutional level, where we have seen a larger embrace of the need to properly resource support for data management. As with any such change, the above institutions can be viewed as early adopters in terms of diffusion of innovations theory (Rogers, 1995). It is hoped that as the national and international research data landscape continues to evolve so the case studies of institutional transformation described above will be followed by the rest of the sector. An ongoing commitment by research funders, as well as the continued support of an organisation such as ANDS, will be critical accelerators of such a transition.

Equally, we are starting to see the positive results of data management and sharing that were hoped for when ANDS commenced activity. ANDS has contributed to these results through carefully-targeted funding of projects and through the provision of the necessary services and support. We expect to continue this support through our institutional engagements, and in collaboration with our international partners, with the goal of enabling all research producing institutions in Australia to become effective supporters of data management.

Acknowledgements

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative.

Thanks to Andrew White, Mat Wyatt, Richard Ferrers and Frances Watson from ANDS for their input.

References

- Australian Government National Health and Medical Research Council. (2007). Australian code for the responsible conduct of research. Retrieved from <http://www.nhmrc.gov.au/guidelines/publications/r39>

¹⁹ Data Management Toolkit: <http://libguides.newcastle.edu.au/content.php?pid=220059&sid=1827507>

- 
- Beitz, A., Dharmawardena, K., & Searle, S. (2012). Research data management strategy and strategic plan 2012-2015. Report for the Research Data Management Subcommittee of the Monash Research Committee, Monash University. Retrieved from <https://confluence-vre.its.monash.edu.au/display/rdmstrategy/Research+Data+Management+Strategy+and+Strategic+Plan+2012-2015>
- Engineering and Physical Sciences Research Council. (n.d.). EPSRC policy framework on research data. Retrieved from <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>
- National Science Foundation. (n.d.). About dissemination and sharing of research results. NSF Data Sharing Policy. Retrieved from <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- Palmer, C. (2012). *Australian Research Council moves to provide better access to data*. The Conversation. Retrieved from <http://theconversation.edu.au/australian-research-council-moves-to-provide-better-access-to-data-9952>
- Pink, C. (2012). The University of Bath roadmap for EPSRC. Research 360. Retrieved from <http://blogs.bath.ac.uk/research360/2012/06/the-university-of-bath-roadmap-for-epsrc/>
- Rogers, E.M. (1995). *Diffusion of innovations*. Fourth edition. New York: The Free Press.