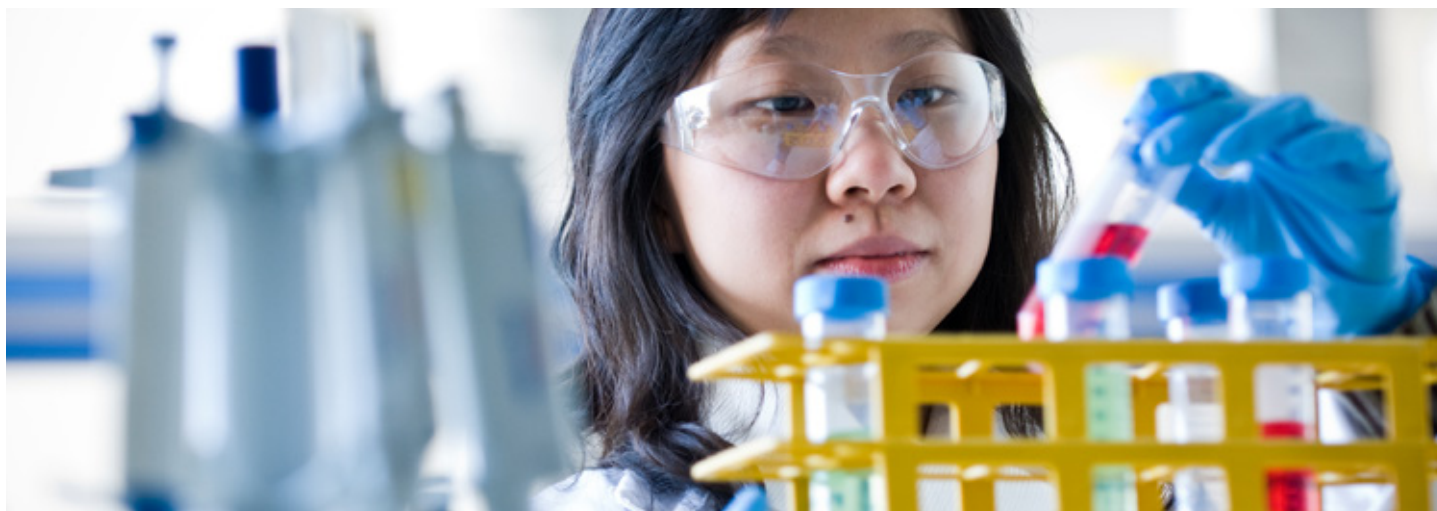


Monash treats research data as an asset

Prof Ian Smith, Vice-Provost (Research and Research Infrastructure), Monash University



Monash values the data its researchers produce, such as from its Bioengineering Laboratory. Image courtesy of Monash University.

Monash University places a high value on the research data produced by its researchers, and recognises this by investing in a variety of strategies to build capacity and capability so data can be properly managed, discoverable, available for reuse and exposed to relevant communities.

The University understands that doing so will contribute to increased research impact, enhanced research practice (including collaboration) and improved educational outcomes, all of which add value to the institution's education and research agenda.

Key to delivering this has been the University's efforts in developing research data management (RDM), using a strategic, multifaceted and multilevel approach.

This approach covers a wide range of external and internal activities, which are intended to deliver value from Monash's research data assets and to foster good research practice.

1. monash.edu/library/researchdata/about/strategy

2. monash.edu/library/researchdata/about/policy

Monash has led and participated in large Australian Federal Government initiatives including VicNode, the Victorian node of the Research Data Storage Infrastructure (RDSI) project; numerous national projects to prototype and develop RDM infrastructure; and is the lead agency of ANDS.

Internally, the University has developed an institutional framework for RDM, developed a Strategic Plan for 2012–2015¹, introduced a RDM policy² with associated procedures and guidelines, and delivered programs for RDM skills and capability building at all levels.

It has operated a multi-petabyte data store since 2006, and continues to develop and deploy a range of discipline-specific and versatile solutions, based on discipline and researcher needs, for the management of research data and associated metadata.

The University continues to identify effective management of its research data assets as critical to its reputation, research

Continued on page 8.

Inside Issue 20 – The value of research data to research institutions

- » Griffith University's engagement with the Gold Coast
- » Building data as an institutional asset
- » Data: White gold for the Australian Antarctic program
- » Institutions use data to attract partners
- » Curtin University and its Major Open Data Collection
- » Data management at the Australian Synchrotron

Griffith University's engagement with the Gold Coast local community

Linda O'Brien, Pro Vice Chancellor (Information Services), Griffith University

Universities play an important role as repositories of existing knowledge, together with a nation's research agencies, libraries and cultural institutions, and are hubs for the generation and exchange of new knowledge.

Crucial to Griffith University's mission is to help drive innovation through collaborating at a number of levels – local, state, national and international – as part of its vision to be a university of influence.



Griffith University is working with local government to ensure the sustainability of the Gold Coast's beaches. Image courtesy of CSIRO (CC BY 3.0)

This is evidenced by Griffith's Centre for Coastal Management (GCCM), which works in partnership with the City of the Gold Coast to improve and develop new best practices in beach sustainability.

GCCM's purpose is to ensure that beaches continue to contribute to coastal lifestyles and Australia's tourism economy into the future, sustainably manage our sandy beach environment and develop coastal protection measures to deal with current erosion issues and forecasted effects as a result of both natural trends and climate change predictions.

At the local level, GCCM has produced a 15-year strategy: *Gold Coast Shoreline Management Plan* (GCSMP). It contains a set of 77 recommendations that will assist the Gold Coast City Council with the progression of best practice coastal management. These recommendations were based on research datasets gathered through a range of major projects.

The GCSMP features recommendations for the economic, social and environmental sustainability of the Gold Coast's beaches. They include actions for capital works and strategies to better understand how beaches function.

While the region benefits from long-term solutions for beach erosion, Griffith University benefits from realising its vision to be a university of influence, adding value to the regions it serves.

Griffith is committed to the importance of Research Data Australia (researchdata.ands.org.au) as a research data discovery service and will be determining how best to release the data that underpins the GCSMP more widely.

Executive Director's report

Ross Wilkinson, ANDS

Prof Peter Rathjen, Vice-Chancellor, University of Tasmania (UTAS), recently addressed a workshop for University Librarians on "Open data to open knowledge: the role of the academic library".

He made the following argument: the reputation of a university is very important. Its library can enhance that reputation, and that is in part based on its collections.

World-class open research data collections, such as UTAS' marine data assets, can help to enhance the university's reputation.

It is a very exciting time for research data in Australia.

Combining the tremendous data assets being generated by the NCRIS (National Collaborative Research Infrastructure Strategy)

capabilities such as AURIN (Australian Urban Research Infrastructure Network), with world-class open data collections such as those Griffith University is building around the Gold Coast, delivers great value to a wide class of researchers in Australia.

However, it is not just researchers who benefit: as Rathjen points out, research data collections can be important to the reputation



Dr Ross Wilkinson

Continued on page 8.

Institutions use data to attract partners

Shannon Lindsay, ANDS

Researchers at Curtin University's John de Laeter Centre for Isotope Research (JDLC) have hit upon a new way to explore diamond deposits in Australia.

The technique uses a Curtin-designed microanalytical instrument that measures the differential diffusion of trace elements in a mineral to determine its time-temperature history.

Last month, JDLC Director Prof Brent McInnes gave a keynote speech at an international conference on thermochronology in France about the team's discovery.

After the keynote, researchers from oil giants Chevron and Shell approached McInnes to find out if this technique could be applied in petroleum system analysis.

"The fact we're producing cutting-edge analytical data does attract industry interest," notes McInnes.

Research data can be influential in gaining institutions key partners for future collaborations.

In the case of Griffith University, these are regional partners, (see article on page 2). For Western Australia's Curtin University, it's the minerals and energy industry. And for the Population Health Research Network (PHRN), it's about attracting other institutions to use and contribute to its research data.

McInnes says both industry and government agencies want to collaborate with Curtin due to the \$25m-plus cutting edge analytical instrumentation it possesses to characterise rock and mineral samples, and the data it produces.

"We work with industry every day, both in a commercial and collaborative sense. The instrumentation capability is a key attractor to industry, but in the long run it's the data from the instruments that industry really wants.

"Industry is attracted to us because we can provide specialist data and information they can't get in-house or through commercial service companies."

With the help of ANDS funding through its Major Open Data Collections program (see article on page 6), JDLC is in the process of making these data publicly available.

"If industry can access the data in a digital form en masse, not only for one area of Western Australia, but entire portions of the country, that could be a powerful exploration tool."

PHRN Chief Executive Dr Merran Smith says its partnerships with research institutions and government agencies increase the richness of PHRN's linked data assets by bringing new data collections into its systems.

"The expanded system, which built on existing capability in WA and NSW, supports a broader range of health and health-related



The mining industry is attracted to Curtin University's rock and mineral data. Image courtesy of Fortescue Metals Group.

research. This in turn contributes to improved health policies and better clinical practice – the ultimate return on the investment in data linkage infrastructure."

PHRN is a research infrastructure network providing safe access to linked administrative data for health outcomes research. The University of Western Australia is lead agent.

PHRN nodes across Australia share data for approved projects with researchers from research institutions such as universities and government agencies.

Institutions with medical, health and population health research programs are major users of the linked data, including the Sax Institute, University of NSW, Curtin University, University of Queensland, University of South Australia, and the Menzies Research Institute Tasmania.

These partners access the data and many also co-invest in PHRN's infrastructure.

One example of PHRN's linked data is its recently completed first 'proof of concept' collaboration on hospital related mortality. This involved linkage of hospital and death records from four Australian states (WA, NSW, QLD and SA), to create a very large research dataset containing more than 45 million records.

For this project, a chief investigator at WA's Department of Health led the researchers, and the PHRN Centre for Data Linkage (CDL) at Curtin University in Perth did the linkage work.

Overall, the research provides new insights into deaths that occur in hospital or soon after discharge. It also describes patterns of illness in people travelling between states or moving interstate to live.

"Prior to PHRN's existence, this type of linkage hadn't been done. You can now get a join-the-dots view across Australia's whole population, whilst preserving privacy in the data," says Smith.

The PHRN is an initiative of the Australian Government, conducted as part of the National Collaborative Research Infrastructure Strategy. "We couldn't have built our systems without their support," says Smith.

Creating value from data: Enhancing institutional assets

Karen Visser, ANDS

Research data are increasingly being seen as a valuable commodity.

Creating value from data: Another brick in the wall by Richard Ferrers and Amir Aryani of ANDS in the last issue of *share* (issue 19, page 6) explored five ways for researchers to capitalise on the value of their data: share data; open data; publish data; talk data; and plan for data.

This article further builds on the concept of the value of data, examining the role of institutions in making data more valuable for researchers.

Managing valuable institutional data assets

Managed research outputs, especially curated and published research data, allow institutions to achieve their mission of accelerating the generation of new knowledge.

Nobel Laureate and Australian National University researcher Prof Brian Schmidt, speaking on the ABC Catalyst program on 4 September this year, talked about the value of sharing data, software, original files and techniques – even with competitors.

"If you want science to go forth quickly, everyone in the world needs to instantly have access to that technique so they check your work out and they can build on your work. That's how science moves quickly."

For data reuse to occur on the scale Prof Schmidt envisages, data cannot be managed and published by individuals or groups of researchers alone. Data management needs the combined services and support at the institutional level.



Prof Brian Schmidt. Image by Belinda Pratten.

Investing in data management will help an institution to track and measure data reuse.

All research institutions are bound by cost-benefit analysis of their business decisions, and the same process of evaluation of benefit from investment is applied across research outputs.

Investments in data management, storage and publication must lead to perceived benefits or 'dividends'.

The 2011 study by John Houghton, *Costs and Benefits of Data Provision: Report to the Australian National Data Service*¹ focused on the costs and benefits to the organisation, the users, and the wider benefits to the economy, separately and collectively. It makes a strong case for increasing the value of research data by freely publishing it. It also demonstrates that the benefits outweigh the costs by some considerable margin.

Importantly, these are not one-off figures, they are annual and ongoing, and so the benefits accrue.

Connectivity

The value of data for institutions is vastly increased through connectivity, which provides context, veracity, metrics and reproducibility. Connectivity also enables a complete demonstration of research outputs.

CSIRO has begun to make connections between data in its Data Access Portal and publications in the CSIRO Research Publications Repository. A good example is the *Time-series for plankton and physio-chemistry of Logan's Dam* dataset which has bidirectional links between the data record² and the publications³, as well as links to the project page which contains the Final Synthesis Report.

CSIRO is also in the process of enabling the publication of software alongside data. The software underlying the *Wheat Scan Images* data collection is available via the Related Materials link⁴.

If data descriptions are wisely connected through rich metadata, they can also provide good reporting mechanisms across a range of research outputs. A good example is the inclusion of Grant Identifiers which can be attached to data records, via PURLs (Persistent Uniform Resource Locators), which are web addresses acting as permanent identifiers that withstand changing web infrastructure.

The Grant ID attached to the Australian Research Council funded project *Boom and bust: the role of fire and rain in driving the dynamics of seeds and rodents in arid Australia*⁵ helped connect and increase the discoverability of a wider range of project outputs, not simply the articles but also including five datasets published in Research Data Australia⁶.

Continued on page 5.

$$\begin{array}{r}
 \text{Benefit} \\
 \hline
 \text{Cost}
 \end{array}
 =
 \begin{array}{r}
 \text{Agency savings} + \text{Users' savings} + \text{Increased returns to annual expenditure on PSI production} \\
 \hline
 \text{Agency costs} + \text{Users' costs}
 \end{array}$$

Human and systems investment in data

Many Australian universities have taken a strategic approach to setting in place systems, policies and investments in their research data over the past decade, and the pace of achievements is accelerating.

Increasingly, we have seen the release of stronger policies and institutional statements aimed at securing and storing research data as an institutional (rather than individual) asset.

Similarly, effort has been devoted to systems development and skill building in the area of research data management and publication.

Monash University's long-term interest and strategic planning in this area is illustrated through its *Research Data Management Strategy and Strategic Plan 2012–2015*⁷, which extends and strengthens statements in the *Summary of research data management at Monash 2006–2011*.

The University of Sydney recently updated its *Research Data Policy 2014*⁸, supported by extensive research data management guidelines documentation⁹.

The guidelines cover managing, sharing, storing, archiving and finding data, as well as relevant policies, ethics, intellectual property and copyright concerns.

Many Australian universities now have comprehensive support materials and services in place to ensure their data assets can be stored and managed – and many of these are in the LibGuide format.

One way to see the impressive range of Australian and international institutional data support services on offer is to simply type into an internet search engine "data management LibGuides".

Similarly, Griffith University has developed a strong partnership between researchers and support and infrastructure services to ensure data are managed and accessible.

Griffith University Pro Vice Chancellor (Information Services) Linda O'Brien noted on the launch of the Griffith research data guidelines: "Data should be an important part of every researcher's strategy to enhance their profile and get their findings out to the widest possible audience. Staff and higher degree students reading the

guidelines will recognise that they are not alone, and that INS [Information Services] has technical solutions and people available to help them capture, store, manage and share their data more effectively."

Ingredients for getting institutional value from research data

Across the globe, we are seeing institutions respond to the growing calls by governments, funders, researchers and publishers to open access to research data for reuse.

Open access, or even mediated access, requires strategic inputs across a range of factors.

Institutions that have begun the process of being able to successfully transform their previously hidden research data into a measurable asset, have some common characteristics:

- » Strong policies which protect and support researchers to publish their data
- » A unified concept that data are an institutional asset, not the responsibility of disparate groups within the institution
- » Connectivity between publications, data, identity, and associated outputs such as software and workflows
- » Systems to harvest metrics associated with data reuse
- » People with skills and comprehensive documentation to support data reuse – encompassing both researcher and support roles.

Publicly funded Australian research institutions needing help with any aspect of research data management are welcome to contact ANDS for advice: contact@ands.org.au

1. ands.org.au/resource/cost-benefit
2. dx.doi.org/10.4225/08/50F62E0D359D5
3. publications.csiro.au/rpr/pub?pid=csiro:EP121781
4. datanet.csiro.au/dap/landingpage?list=BRO&pid=csiro:8827&sb=RECENT&rn=6&rpp=25&p=1&tr=37&bKey=td&bVal=Biological%20Sciences&dr=all
5. purl.org/au-research/grants/arc/DP0451749
6. researchdata.ands.org.au/boom-bust-role-arid-australia/62572
7. monash.edu/library/researchdata/about/strategy
8. sydney.edu.au/policies/showdoc.aspx?recnum=PDOC2013/337
9. sydney.edu.au/research_support/data/

Curtin creates digital mineral library: ANDS Major Open Data Collections

Shannon Lindsay, ANDS

Sitting on a shelf in a Perth warehouse, largely inaccessible to the public, are up to 2,000 heavy mineral concentrate samples from the Geological Survey of Western Australia (GSWA).

For more than 30 years, government geologists have systematically collected samples from key rock formations in order to determine the geochronological age and evolution of WA using the analytical instruments in Curtin University's John de Laeter Centre for Isotope Research (JDLC).

Mindful of the continual advances in technologies, the GSWA archived the residual sample material for future generations of researchers to add value using better analytical tools and techniques.

As JDLC Director Prof Brent McInnes puts it, unless the research community is aware of their existence, these samples are "stranded assets".

That is why Curtin and GSWA have joined forces to create a digital mineral library of the collection with support from the ANDS Major Open Data Collections (MODC) program launched this year.

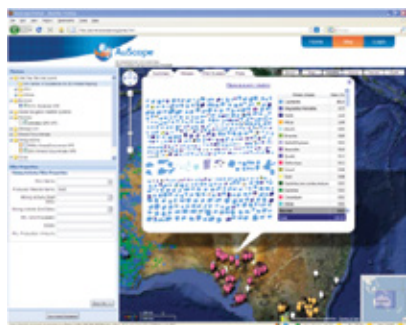
This project will use state-of-the-art technology to produce mineral maps of the samples, and publish these "digital" mineral collections via the AuScope Discovery Portal (portal.auscope.org), a national earth science infrastructure program.

This ANDS-funded MODC project will create an appropriate metadata schema for these datasets available via Research Data Australia (researchdata.ands.org.au) and the AuScope portal to facilitate discovery and access to the sample information by the national and international research community.

For GSWA, the JDLC digital mineral library will allow it to better understand the scientific value of its collection, and assess sample requests from collaborating research organisations in a structured way.

JDLC software engineer Adam Brown, who is designing the Laboratory Information Management System (LIMS) for the project, says there are several advantages in making these datasets publicly accessible and reusable.

"Making this collection and its associated datasets discoverable



A mineral map on the AuScope Discovery Portal. Image courtesy of JDLC.

will increase the uptake of data generated in JDLC facilities, and Curtin as the Centre's agent will obtain greater recognition through higher data citation rates," says Brown.

"It is also likely that people from other institutions or companies will search the database and contact us about collaborative research on the samples."

McInnes also sees opportunities in the data management infrastructure itself. He was recently at an international conference and laboratory data management processes was a hot topic.

"Everyone has the same issue in that funding agencies are increasingly requiring researchers to make their publicly funded datasets more accessible to the public. I told the audience about what we're doing in our MODC project, and a lot of people asked for further information.

"We're further advanced in our thinking on data management practices than our counterparts in leading research institutions, such as UCLA, Stanford and CalTech – however they don't have the advantage of having a data portal already constructed like we have with the AuScope Discovery Portal.

"Once we demonstrate we have a solution to an important problem faced by thousands of researchers at hundreds of universities around the world, there could be lots of opportunities for collaboration," says McInnes.

Curtin might not be the only Australian institution positively in the spotlight in the future. ANDS is providing funds and expertise to more than 20 publicly funded research institutions across Australia as part of the MODC program, running from 2014 to mid-2015.

The aim of ANDS' MODC program is to develop internationally significant collections that provide institutions with a valuable opportunity to enhance their research data assets, enable new partnerships to be formed, and enable new data intensive approaches that address internationally significant research challenges. (For more information and to view other MODC projects, visit the ANDS Projects webpage¹.)

Curtin University, for one, is already thinking about taking its MODC project even further.

By working with AuScope and industry, McInnes says his team hopes to turn the digital mineral library into a "truly national project" by incorporating more than 300,000 mineral samples from across Australia.

1. projects.ands.org.au/getAllProjects.php?start=modc

Data: White gold for the Australian Antarctic program

Ben Raymond and Dave Connell, Australian Antarctic Data Centre

Collecting scientific data from the Antarctic and Southern Ocean is expensive and difficult because of the harsh and remote environment.

But these data are incredibly valuable, supporting climate science, ecology, oceanography, and other research, as well as conservation management actions such as the development of marine protected areas.

Partly for this reason, the Antarctic science community has a long-standing history of collaboration and data sharing.

The 1959 Antarctic Treaty, to which Australia was one of 12 original signatory countries, declares "scientific observations and results from Antarctica shall be exchanged and made freely available".

This treaty obligation is reflected in our Australian Antarctic program, which has made significant financial and intellectual investments in data-related matters.

All data and samples collected under the program remain the property of the Commonwealth of Australia, and investigators must ensure that data are adequately managed for long-term reuse.

The Australian Antarctic Data Centre was established in 1996 as the primary facility for this purpose, and currently lists 2,300 records in its metadata catalogue. These records are syndicated into Research Data Australia¹ and by other repositories internationally.

The Australian Antarctic program's data policy² describes the responsibilities of scientists collecting data, and of the Data Centre.

Compliance with this policy is considered when new project proposals are assessed. All projects are required to develop a data management plan, which provides an outline of the data to be collected and the associated action needed to manage the data.

This data culture is supported by the Australian Antarctic Division. Each project has a designated Data Centre staff member to act as a liaison on data-related matters. The Data Centre also has a staff member whose primary role is to archive and catalogue data from program participants. This ensures that scientists have the support they need to fulfil their obligations, as well as ensuring that our metadata and data meet expected standards.

Open access brings many benefits, such as reusing existing data and producing new syntheses of information. A recent and striking example of this is the *Biogeographic Atlas of the Southern Ocean*: a 500-page book written by 150 contributors across 22 countries, based on more than one million data points (species observations) collected by the international Antarctic research community over many decades³.



Scientists collecting sea ice cores from the Antarctic. Image courtesy of the Australian Antarctic Division.

Open access also allows new data acquisition activities to be designed to best complement existing data, perhaps by filling gaps in existing knowledge.

Sometimes data can prove to be useful in ways that were not anticipated at the time of collection. For example, locations of whaling catches in the early 20th century have been used to infer the historical extent of sea ice, since whaling activities were concentrated near the ice edge⁴.

An enduring problem has been quantifying the use and reuse of data, so that the efforts of data contributors can be adequately recognised.

Some datasets have been published as data papers in scientific journals, which bring the advantages of a formal citation and a Digital Object Identifier (DOI) on the paper that can be tracked over time.

In addition, the Data Centre and ANDS have been issuing DOIs on datasets themselves, providing formal identifiers and citations for those datasets.

Scientists are encouraged to self-cite their own data, to help establish formal data citation as a standard practice in our program.

The Australian Antarctic Data Centre will continue to invest in and improve our data culture, in order to support our science and strengthen its impact, and to take advantage of emerging opportunities in an increasingly data-centric world.

1. researchdata.ands.org.au/search/#/1/class=collection/group=Australian%20Antarctic%20Data%20Centre/

2. data.aad.gov.au/aadc/about/data_policy.cfm

3. De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C d', Van de Putte AP, Danis B, David B, Grant S, Gutt J, Held C, Hosie G, Huetmann F, Post A, Robert-Coudert Y (eds.), 2014. *Biogeographic Atlas of the Southern Ocean*. Scientific Committee on Antarctic Research, Cambridge, xii + 498 pp. ISBN: 978-0-948277-28-3

4. de la Mare WK (1997) *Abrupt mid-twentieth-century decline in Antarctic sea-ice extent from whaling records*. *Nature* **389**:57–60

Monash treats research data as an asset

(continued from page 1)

performance and to the fulfilment of compliance requirements and community expectations.

At Monash, all members of the institution share responsibility for undertaking and enabling RDM, in a coordinated and integrated way.

To facilitate this, the University has active internal collaborations between the Library, Monash eResearch Centre, Research Office, eSolutions (the University's IT services unit) and all faculties, and has made appointments and secondments into research data management roles to build capability and expertise.

This pool of expertise has a shared focus on common outcomes, which will drive Monash forward as researcher needs, university goals and technology evolve.

Underpinning this shared focus are formal, documented RDM policies and strategies; established organisational structures to plan, implement and promote RDM; a RDM website³, that provides guidelines, training materials and other resources; programs to build capability in both professional and research staff; established technical infrastructure that encompasses a number of RDM platforms designed to meet different user needs; a large, well-used and widely accessible institutional digital data store; and the Monash University Research Repository⁴ to draw the descriptive metadata about the University's research together and disseminate it to the world.

This flexible research infrastructure provides Monash with the tools it needs for the future.

In strategically significant areas Monash is providing discipline-specific RDM platforms, such as Interferome⁵, which assimilates a large number of datasets, including detailed annotation and quantitative data, from the microarray analysis pipeline, and MyTARDIS⁶, an automated solution for managing and sharing raw protein crystallography data.

For the broader University community it is providing a variety of versatile RDM solutions. These technical solutions allow the University's research data assets to be described, managed, shared, published, discovered and reused.

All have access to advice and ongoing conversations on the process of curation and exposure of research data collections.

The sophisticated infrastructure now in place at Monash represents critical steps towards maximising the value of the institution's research. This will improve research practice, attract researchers and research income, encourage reuse of data to increase research impact, validate research outcomes and reduce risk.

3. monash.edu/library/researchdata

4. arrow.monash.edu

5. <http://interferome.its.monash.edu.au/interferome/home.jsp>

6. <http://mytardis.org/>

Executive Director's report

(continued from page 2)

of research institutions, but also very important in enabling institutions to drive their research strategies and partnerships.

We have previously seen in *share* descriptions of how James Cook University is using its Tropical Data Hub (tropicaldatahub.org) to drive its tropical research strategies and partnerships.

Many other institutions are taking this strategic approach as well.

Monash University is assembling open data collections from the unique array of imaging facilities located around the university, ranging from microscopes to the Australian Synchrotron (see story on page 9).

As mentioned, UTAS is establishing a major open marine data collection that is strengthened by the Integrated Marine Observing System (imos.org.au), a national approach to assembling marine data from multiple ocean observing instruments.

Prof Craig Johnson, Director, Marine and Antarctic Futures

Centre, observes that these data assets already have strengthened UTAS' international partnerships, as well as attracting students from around the world to study at UTAS' Institute for Marine and Antarctic Studies.

Internationally, research data are being taken increasingly seriously. For instance, the data being generated out of the European Commission's €80bn Horizon 2020 research program are expected to be made publicly available.

We are going to see even more open data internationally.

Australia's research institutions have a wonderful opportunity to position themselves as custodians of some of the most important research collections. Collections that are long-lived, collected with purpose, and well managed can be an enduring asset of Australian research institutions seeking to partner with the best on global research challenges.

Breaking down the boundaries to storing, sharing and publishing research data

Steve Androulakis, Philip Bertling, David Groenewegen and Andrew Harrison, Monash University¹

For several years, Monash University has employed the Data Curation Continuum² as a conceptual model to understand the nature of research data workflows, and to put in place infrastructure that will enhance and enable these workflows.

The Continuum separates the process of research data management into three distinct domains: Private, Shared and Public.

These domains are separated by "curation boundaries", which are virtual decision points at which the creators of data decide what they will share, with whom, with what metadata and under what conditions.

Traditionally, these boundaries have been crossed using some form of manual intervention – either data were moved from one repository to another that is more open, or metadata was created only at the point where it was needed to cross the boundary. This process was generally time-consuming and inefficient.

Store.Synchrotron (store.synchrotron.org.au) is a service established at the Australian Synchrotron in partnership with Monash University to assist the management, access and sharing of the raw x-ray diffraction data produced from two of its instruments.

X-ray diffraction is a technique used by protein crystallographers in order to determine the atomic and molecular structure of a protein, and hence its biological function.

Via Store.Synchrotron, data are automatically stored in an organised and detailed manner browseable on the web, reducing the need for researchers to download it all locally in order to filter and sort their results.

The same web interface enables researchers themselves to share data securely with collaborators and the public. And the publication process results in a browseable data page and a Digital Object Identifier (DOI) for citation in research publications.



The Australian Synchrotron. Image courtesy of Monash University.

Prior to the establishment of Store.Synchrotron, access to the hundreds of gigabytes acquired by a researcher once they return to their institution was a difficult and risky task requiring physically moving portable hard drives. Subsequent data sharing with collaborators, particularly internationally, was difficult and time-consuming.

Store.Synchrotron facilitates the secure sharing of data with users nominated via the Australian Synchrotron's identity system, the Virtual BeamLine (VBL). Collaborators are able to login to the same system, browse and download raw data.

Connecting data in this way eliminates the need for it to be transported via physical storage media.

The publication of raw research data online is often extremely difficult due to the financial and technical difficulties of sharing large amounts of data on the web. A self-serve publication module has been developed to assist the researcher in describing their raw data with a high level of detail before releasing it to the public.

Given the potentially wide audience for high-impact research, it's important to provide an adequate level of general descriptive metadata. This includes information such as the creator(s), date of acquisition, kind of data produced and how one might view and process it.

Store.Synchrotron benefits from the early capture of data, with several descriptive elements from the Australian Synchrotron's internal databases automatically included at the time of data generation. The process also assists in the collection and capture of discipline specific metadata.

For the collecting researcher, the Store.Synchrotron service enables the archiving of data early and automatically (the Private Domain); facilitates controlled collaboration through tools provided (the Shared Domain); and provides a mechanism to make data publishable and freely available (the Public Domain) without the researcher needing to shift data from one place to another.

The service operates in a cloud that is physically close to the data and is designed to withstand the increased network and computational burden of increased data and usage.

1. For more detail see: Androulakis, Steve; Bertling, Philip; Groenewegen, David; Harrison, Andrew, *Breaking down the boundaries to storing, sharing and publishing research data*. Presentation at *Open Repositories 2014*, Helsinki, Finland, June 9-13, 2014
urn.fi/URN:NBN:fi-fe2014070432369
Meyer, G.R., Aragao, D., Mudie, N.J., Caradoc-Davies, T.T., McGowan, S., Bertling, P.J., Groenewegen, D., Quenette, S.M., Bond, C.S., Buckle, A.M., Androulakis, S. *Operation of the Australian Store.Synchrotron for macromolecular crystallography* (2014). Acta Cryst. D70, doi:10.1107/S1399004714016174.

2. ands.org.au/guides/curation.continuum

Chair's report

Ron Sandland, ANDS' Steering Committee Chair

Following the successful funding of NCRIS 2015, work is now in full swing to finalise the plans of the capabilities, including ANDS, and also to prepare the arguments for continued funding of NCRIS (National Collaborative Research Infrastructure Strategy) beyond 2015.

In order to accomplish these objectives, the Australian Government has decided to conduct a number of reviews to understand how the investment process might provide the most effective outcomes and the best value for money.

Emeritus Prof Tom Cochrane, Queensland University of Technology, is undertaking a formal review of the status of the national eResearch infrastructure capability. This report will address a number of key questions about the appropriateness of the current investments, their contributions, governance, lessons learned and constraints.

A parallel review will be carried out by Dr Rhys Francis, Director, eResearch Futures, to develop a strategic framework and architecture for future national eResearch infrastructure to meet the needs of the research sector to 2020.

Along with these important reviews there will also be a survey and an efficiency review.

One of the challenges the Department of Education, which oversees NCRIS, will face in drawing all this information together is the intrinsic differences between the various capabilities that constitute NCRIS.

These differences have not arisen simply because of different approaches taken by the capabilities; much more importantly they have arisen because of the nature of the issues they have been dealing with.

The philosophy of ANDS has always been to work with institutions. And the results of that philosophy, and the hard work put in to enact it, have been transformative.

ANDS has established strong relationships with almost all of Australia's research institutions through such initiatives as establishing the Australian Research Data Commons populated with more than 100,000 research data collections.

Along the way it has dramatically improved institutional research

data management and helped significantly in the establishment of institutional research data infrastructure.

This institutional capability is arguably the best in the world. It was as a result of this coherent capability that Australia was invited to take a leading role in establishing the international Research Data Alliance (rd-alliance.org).

This acknowledgement and the capability that underpins it means that Australian researchers are very well-placed as collaborators of choice. Effective global collaboration is key to the recognition of Australia's researchers and their institutions.

But to play the devil's advocate for a moment, what would be different if all this hadn't happened? Wouldn't Australian institutions have recognised this need and acted on it anyway?

Unfortunately the more likely outcome would have been a plethora of uncommunicative capabilities that might have met the internal needs of institutions, without enhancing collaboration between institutions. It would have resulted in a significant reduction in the efficiency and effectiveness of collaborative data-intensive Australian research.

These ideas were well explored in the development of TARDIS (The Australian Research Data Infrastructure Strategy), which is now safely in the hands of the Department of Education. Two critical recommendations were:

- » Design national research data infrastructure investments to enhance and complement capability at the institutional and international level, encourage institutional commitment, and foster sustained institutional capacity in data-intensive research.
- » Coordinate research data infrastructure investments to drive collaboration between and commitment from research institutions, government agencies, industry participants and community stakeholders.

These recommendations are perfectly aligned with the mission of ANDS, as a result of which we are a long way down the path of realising them.

Sharing sensitive data

Shannon Lindsay, ANDS

ANDS has been on a sensitive data blitz over the last few months, producing a new guide and webpages, and developing webinars and presentations on this tricky topic.

Key amongst ANDS' new sensitive data material is the guide, *Publishing and Sharing Sensitive Data* (ands.org.au/datamanagement/sensitivedata).

The guide, authored by ANDS' Dr Sarah Olesen, provides straightforward advice about what you need to know and do before publishing and sharing sensitive data.

Olesen will present a paper at the eResearch Australasia conference in Melbourne on Tuesday, 28 October on sharing sensitive data safely and simply.

She has also organised two sensitive data webinars via ANDS. Check the ANDS Events webpage for further information (ands.org.au/events). If after the fact, check the 'andsdata' YouTube channel for the recorded webinars (youtube.com/user/andsdata).

Research Data Alliance: Reaping the fruits in Amsterdam

Andrew Treloar, ANDS

At the Research Data Alliance Fourth Plenary Meeting, the research community was urged to change its culture around research data and treat them "as you treat your publications".

The entreaty came from Robert-Jan Smits, General Director, Directorate-General for Research and Innovation Research and Innovation, European Commission, in his opening keynote.

The plenary was held in Amsterdam from 24–26 September 2014.

The Research Data Alliance (RDA) is a rapidly growing international organisation (from eight to 2,300 members in just 18 months), with a vision of a world where researchers and innovators openly share data across technologies, disciplines, and countries to address the big challenges of society.

The RDA accomplishes its mission through Working Groups (producing deliverables within an 18-month time frame that will be implemented and adopted by one or more specific communities), and Interest Groups (serving as a platform for communication and coordination among individuals with shared interests).

The title chosen for the Amsterdam event was "Reaping the Fruits" – a reference to the opportunity to showcase the outputs of the first Working Groups.

The research funding organisations that sponsored the creation of the RDA have invested in it because they want more cost-effective research infrastructure.

And so the main themes discussed at this event related to the first fruits: working to get adoption of RDA outputs, and ensuring they are as adoptable as possible.

From an Australian point of view, this was a very successful event. There were 12 attendees from Australia, drawn from ANDS, CSIRO, AURIN (Australian Urban Research Infrastructure Network), TERN (Terrestrial Ecosystem Research Network), NICTA (National Information Communications Technology Australia), and the University of Queensland.

The Australians held a Birds of a Feather session on sustainability of eResearch outputs, led a Working Group on Data Description Registry Interoperability, and co-led Working Groups on Data Publishing Services, and Urban Quality of Life Indicators. And of course took part in more than 70 working sessions at the plenary and contributed their efforts to Council, the Technical Advisory Board and Secretariat.

The most significant changes between the Third and Fourth Plenaries were both quantitative and qualitative.

The numbers of RDA members, attendees and Working and Interest groups were all up.

The quality of the discussions at group working sessions was greatly improved on previous plenary meetings.

But most importantly, the RDA community is clearly maturing in its understanding of the role it can play in delivering the RDA vision.



Members of the Research Data Alliance Council at the Fourth Plenary, including ANDS Executive Director Dr Ross Wilkinson (front, green shirt), and ANDS Technology Director Dr Andrew Treloar (standing, third from left).

ANDS at eResearch Australasia 2014

Workshops

Creating your research impact story

Presenters: Heather Piwowar (Impactstory) and Karen Visser (ANDS)

When: Monday, 27 October, 9am–12:30pm

Humanities data

Presenter: Ingrid Mason (ANDS)

When: Friday, 31 October, 9am–5pm

Presentations

I'm a Sensitive Soul: How to share sensitive data safely and simply

Presenter: Sarah Olesen (ANDS)

When: Tuesday, 28 October, 3pm–3:30pm

The ResearchLink Initiative: The nexus between data, publications, and grants

Presenters: Adrian Burton and Amir Aryani (ANDS)

When: Tuesday, 28 October, 5pm–5:30pm

The Value of Research Data to the Nation

Presenter: Richard Ferrers (ANDS)

When: Wednesday, 29 October, 12:30pm–1pm

Birds of a Feather

The Research Data Alliance

Presenters: Stefanie Kethers, Andrew Treloar and Amir Aryani (ANDS), alongside Mark Parsons (Rensselaer Polytechnic Institute, Troy, NY) and Simon Cox (CSIRO)

When: Tuesday, 28 October, 2pm–3pm

rd-alliance.org

Full program and register: conference.eresearch.edu.au

Come and say hello to the ANDS team at the ANDS booth!

Follow the conference via Twitter: [#eRes2014](https://twitter.com/eRes2014)

Follow ANDS via Twitter: [@andsdata](https://twitter.com/andsdata)



Heather Piwowar coming to Australia

Heather Piwowar, co-founder of Impactstory (impactstory.org) and a global leader and researcher in research data availability and reuse, will feature in a series of workshops in three Australian state capitals. Visiting Australia from the University of British Columbia, she wrote one of the first papers measuring the citation benefit of publicly available research data¹. She has studied patterns in data archiving and data reuse, and the impact of journal data sharing policies.

Opportunities to talk with Dr Piwowar while she is in Australia are as follows:

- » Melbourne, Mon, 27 Oct, 9am–12.30pm: 'Creating your research impact story': Workshop at eResearch Australasia. Also with Pat Loria, Charles Sturt University, and Natasha Simons, ANDS.
- » Melbourne, Wed, 29 Oct, 10–11am: Keynote presentation at eResearch Australasia.
- » Brisbane, Mon, 3 Nov, 1–4.30pm: 'Uncovering the Full Impact Story of Open Research'. Also with Paula Callan, Queensland University of Technology, and Ginny Barbour, PLoS.
- » Sydney, Wed, 5 Nov, 1.30–4.30pm: An afternoon of talks featuring Dr Piwowar, Maude Frances, University of NSW, and Susan Robbins, University of Western Sydney, at the Australian Catholic University.

See the ANDS Events webpage (ands.org.au/events) for further information and to register for these events.

1. [http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000308?utm_source=andsUP&utm_campaign=b84124cfd-andsUP_30_September_2014%202014&utm_medium=email&utm_term=0_22b849a4ee-b84124cfd-16191493&ct=t\[andsUP_24_June_2014%2010_2014\]](http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000308?utm_source=andsUP&utm_campaign=b84124cfd-andsUP_30_September_2014%202014&utm_medium=email&utm_term=0_22b849a4ee-b84124cfd-16191493&ct=t[andsUP_24_June_2014%2010_2014])

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy program.

This newsletter is designed by GRIT Brand + Online (gritcomms.com.au)

ANDS Project Partners:



MONASH
University



Australian National University



NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative



With the exception of logos or where otherwise indicated, this work is licensed under the [Creative Commons Attribution 3.0 Licence](http://creativecommons.org/licenses/by/3.0/).