

Discovering our Research Data

Ross Wilkinson, ANDS



Discovering the right research data can be crucial to research. It is a lot easier than having to create it from scratch, or even worse, re-creating it. It may also add data from other times or places to augment the new research.

Discovery is tackled in many different ways. Google's web search succeeded because it provided a simple way of finding a web page where previous search engines had tried to answer questions. The path to discovery often takes one via a circuitous route. For instance, supermarkets provide navigation to the milk but the path may also take you past the chocolate, while online versions will take you past customized offerings and advertisements. Similarly, we often find music by getting recommendations from friends or on-line sites. Navigating your way through the deluge of research data is just as complex.

In the previous issue of *share*, the Australian National Data Service (ANDS) described the rich connections that can be made between research data collections and their context, and all of these connections are important in discovery. ANDS supports discovery of research data through Research Data Australia (<http://services.ands.org.au>), its window into the Australian Research Data Commons. This website enables simple text search, navigation through the connections and spatial filtering. Each collection description and

its associated entities are published separately as a web page, so they can also be found with a web search—try typing "Mawson research data" into your favourite search engine.

A very important way of discovering research data is by way of citation. Following a link from the citation to the service that resolves that citation to a destination, means that research data can be discovered in the same way that a journal paper can be discovered.

Portals are another way of enabling discovery. Many disciplines are supported by specialist portals such as the Australian Ocean Data Network's portal (<http://portal.aodn.org.au>) which is tailored to discover marine data or the Atlas of Living Australia's portal (<http://ala.org.au/>) which is tailored to biodiversity information. Some institutions are providing access to their own research data; for example, CSIRO's Data Access Portal (<http://datanet.csiro.au>) and Griffith University's Research Data Management Centre (<http://equella.rcs.griffith.edu.au/research/access/home.do>).

Richly connected data is more easily discoverable, whether via a simple search, navigation or specifically tailored tools. All of these discovery methods can lead to the data you need to enhance your research.

Inside Issue 10 – Special Issue on Discovering Research Data

- » Discovering our Research Data
- » ANDS out and about—report on completed ANDS-funded projects
- » Chair's report—Ron Sandland
- » ANDS event reports
- » Conference reports
- » Meet the ANDS staff
- » In brief
- » Forthcoming events

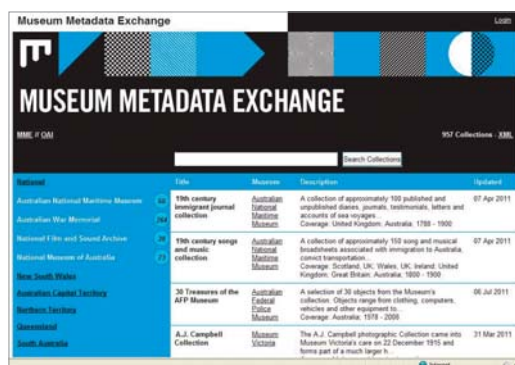
Unlocking the power of specialist collections

Tim Thwaites, Science in Public

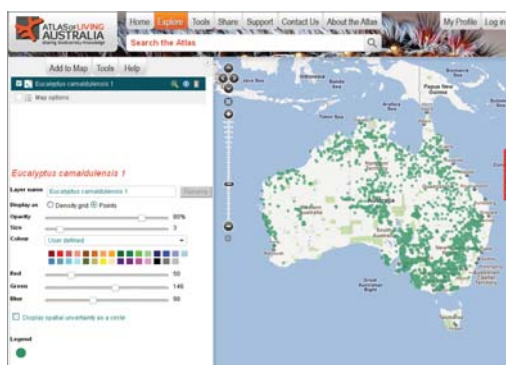
How did ordinary people dress in the 1920s? In standard reference books, the clothing you find tends to be fashion worn by the very rich. It would be like using Vogue magazine as the reference guide for what we're wearing in 2011.

But there is a source of information on the street clothes of the time which is absolutely authentic, if unexpected—crime scene photos. That's why cultural researchers ended up trawling through the archives of the Australian Justice and Police Museum at Circular Quay, which has the world's largest collection of crime scene photographs of the 1920s.

In the past it might have taken creative insight, personal knowledge or word of mouth for researchers to find such a treasure trove of information. In future, they should be able to discover such valuable resources by electronically searching the Museum Metadata Exchange (MME), a database which will provide researchers in the humanities and social sciences (HASS) disciplines with a snapshot of the information housed within Australia's museums that may be of relevance to their specific research projects.



Access to hundreds of collections from museums all over Australia is provided through the MME portal



Atlas of Living Australia, an example of data discovery for the Red River Gum. Image courtesy of Robyn Lawrence, CSIRO

To generate its backbone, 14 significant national, state and regional museums were asked to submit information on about 50 social, cultural, historical or ethnographic collections in their keeping. This was not as simple as it sounds.

"It was a first for many museums," Anderson says. "They're used to working with objects. Everything in a museum is catalogued as a distinct item." But Research Data Australia wanted descriptions of objects grouped together in ways that would be helpful to answer research questions—anything to do with alcohol use in NSW before 1850 or house design in the outback.

The task made museum curators look at their holdings in a new way and already, the museums involved have developed a new understanding of what they hold and how it fits in with the collections of others. Much that was hidden has now been exposed, in a completely new light.

A more developed activity is the Atlas of Living Australia (<http://www.ala.org.au/>), which provides links to the huge variety of information on the organisms inhabiting Australia. For the River

Red Gum, for instance, the Atlas can direct you where to find information on its classification and identification, its biology and ecology, its distribution, and its environmental needs. The Atlas already has associated tools which map the distributions of species, allow you to explore what species can be found in your local area or region, help identify species, and even allow you to submit information of your own.

The MME (<http://www.powerhousemuseum.com/museumexchange/>), which is still under construction, is one of an increasing number of discipline-level portals into Australia's rich accumulation of collected information. They include the Australian Ocean Data Network (AODN), the Atlas of Living Australia (ALA), the AuScope Discovery Portal for earth sciences, and the data services of the European Bioinformatics Institute at European Molecular Biology Laboratory (EMBL) Australia.

"Until now there hasn't been any attempt to explore the potential of the humanities and social sciences collections in museums," says Margaret Anderson, Chair, Council of Australian Museum Directors and Co-Chair, MME Steering Committee. This project, funded by ANDS, also created a metadata store that feeds automatically into Research Data Australia: the ANDS data portal.

Access to such information was identified as a national infrastructure need and the project has received \$38 million from the Federal Government since 2007. "The interactions between agriculture and other development activities, and native environments are very important for the Federal Government, the state governments, all the way down to catchment and local governments and natural resource management groups," says Donald Hobern, the Director of ALA.

Because ANDS is working to link together these different discipline collections, soon it will be possible to move from an old photo of a platypus-skin cloak, to information about current platypus habitat, to detailed data about its genome. As the number of collections grows, the potential for serendipitous discovery will also increase, allowing more researchers to discover the value of cross-disciplinary data.

Discovering rich data that answers questions

Tim Thwaites, Science in Public

Someone somewhere has data that could help your research. But do you know where to look? Research Data Australia will search for you, and find data collected by dozens of Australian research organisations. Until you try it you'll have no idea what serendipitous connections you could be missing.

Insert the word "Mawson" into the Research Data Australia (RDA) search engine (<http://services.ands.org.au>) and up pops a list of 140 datasets, almost all to do with Australia's Antarctic base of that name. The topics they cover range from ocean salinity and sea surface temperatures to tourist numbers and seabird strikes on equipment.

You could find many of the same scientific observations by going directly to institutional websites, such as the Bureau of Meteorology, or discipline collections, such as the Australian Antarctic Data Centre. But you would be unlikely to discover information on the base founder Philip Law or the Russian observations of sea ice thickness.

These are all captured, however, in Research Data Australia: the Australian National Data Service's portal. What's more, each dataset is accompanied by information on the people who gathered the data and their research activity, the techniques they used, the context in which the data was collected, links to the organisations for which they worked, and more.

After traversing this information mesh, and finding something you want, RDA will refer you back to the data centres, such as the Bureau of Meteorology or the Australian Antarctic Data Centre, to access the data.

The RDA pages describing datasets are written in a way so they are accessible to general search engines, such as Google, Yahoo! and Bing allowing the datasets to be accessed directly through the web. Google "Mawson salinity data", for instance and you will find that the RDA pointers to this data are displayed prominently in the search results.

"We think of Research Data Australia as a rich data discovery service," says Dr Ross Wilkinson, Executive Director of the Australian National Data Service. "You're finding much more than just the data. You're finding its entire context—and that can be really important."

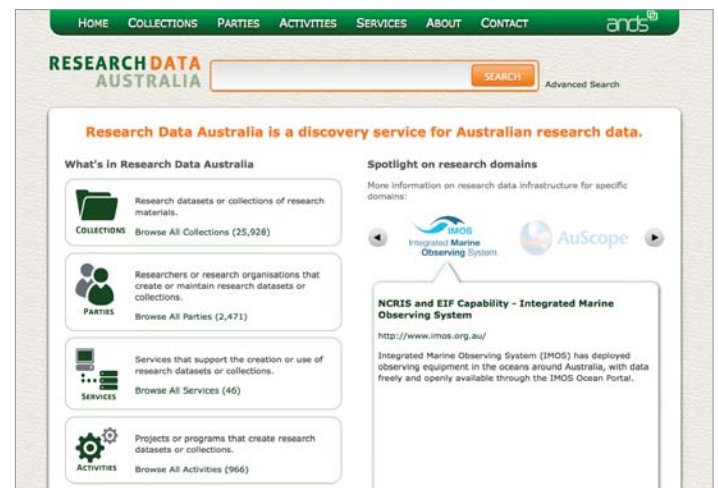
Because of the investment in RDA, Ross says, "Australia is now amongst the best in the world in creating national level research data infrastructure." This gives researchers a capacity to pose and solve many new research questions on a national scale, such as how Australian society adapts to climate change or responds to the obesity epidemic.

"There's a fantastic project called Founders and Survivors where the integration of convict records and First and Second World War medical records enables you to do very long term epidemiological studies across families." To put the collection together, researchers from five universities needed linkages, which came out of the work of genealogists. RDA enables increased discovery of this data.

"The key to Research Data Australia is that government is increasingly

willing to invest in research tackling large challenges. You need appropriate infrastructure to do that. It's no different from acquiring a new instrument. In the same way that building a synchrotron allows you to probe bigger, deeper problems involving materials, so does access to a larger pool of data enable you to answer new questions."

The development of RDA is based on initial work undertaken at the Australian National University by the Australian Partnership for Sustainable Repositories. But it is much more, says Ross Wilkinson, than the mechanics of building a list of collections and some fancy software to search it.



Home page of Research Data Australia

For a start, the owners of the data need to see the value of making their resources available—the idea that data can be valuable beyond the purposes for which it was gathered. Then they need to agree on how to make their resources discoverable and what the protocols are. "You need to provide a generic search capability, which is not specific to any particular discipline, so that a geographer can access data from a historian or a geologist or an astronomer."

"Research Data Australia is not only a discovery portal, but also a publicity portal. It's a way to make research data visible to others." Even though RDA is already usable and covers all major research areas, it is still very much under development in terms of its power and the amount of data it reaches.

"Research Data Australia is providing access to information that we're harvesting from around the country. Most universities have automatic feeds. As the institution adds data to its system, we provide access to those collections." And the same is true for government research organisations, such as Geosciences Australia or CSIRO, and non-government research institutions.

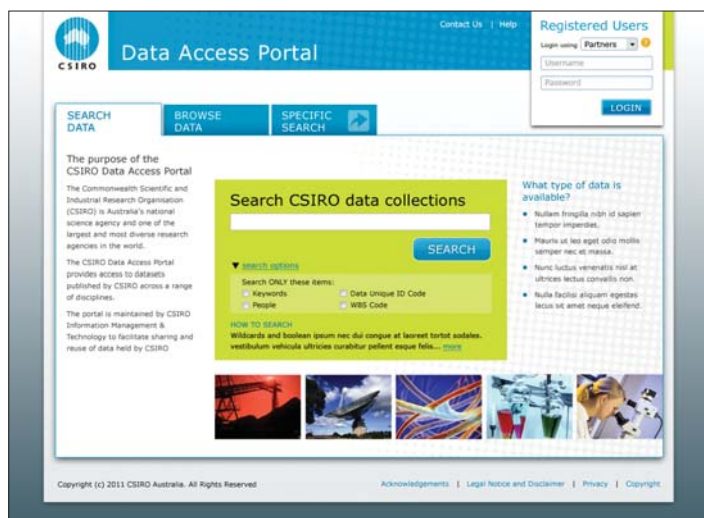
Discussions are already under way with European and American authorities, Ross says, on how Research Data Australia can be integrated into international data discovery systems.

Exploring the breadth of an institution's data

Cynthia Love, ANDS and Tim Thwaites, Science in Public

CSIRO's radio telescope at Parkes is used by astronomers internationally and has been involved in the discovery of more pulsating neutron stars or pulsars than any other facility. Data captured from this facility is likewise of interest internationally. In a recent project funded by ANDS this data was brought together, secured, stored, and made accessible to other researchers from a single site. The results already have been striking, says Dan Miller, the Project Manager. "Researchers in China are already using the accumulated data and new papers have been written," he says.

This work has formed the basis for the development of an enterprise-wide Research Data Service and the technology to support it. "There has been a growing awareness of research data management within CSIRO," says David Toll, Chief Information Officer of CSIRO. "Key benefits to CSIRO in building research data management capability include improved access, reuse and increased citation of data, as well as securing valuable research data for the future."

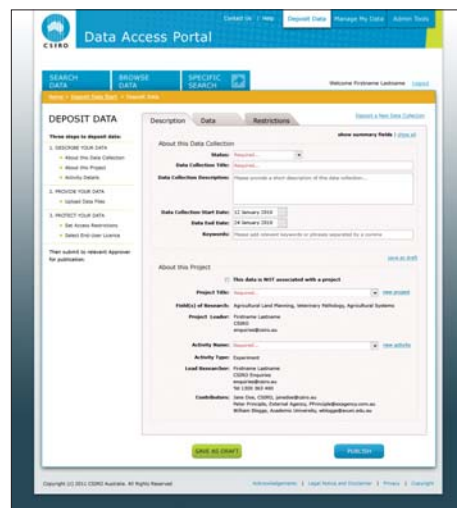


An example of the Data Access Portal deposit data interface

The system, with its Data Access Portal (<http://datanet.csiro.au>), provides a management and discovery tool for the breadth of CSIRO's data, which ranges widely across scientific research. In order to provide this comprehensive view and remain relevant to the research, the development team had to identify the commonalities and then the points of difference in the data, the descriptions and associated access.

Miller says, "We are developing a baseline platform, which has a framework of generic tools—tasks such as depositing and describing data, supporting workflow for approving the access and publication of data, and providing search functions for discovery and reuse. On top of the base platform, we are building domain specific extensions to provide additional functionality for specific research communities."

The Data Access Portal will provide everyone, inside and outside of the organisation, with information on what data CSIRO holds. It delivers both browse and search functionality, so users can simply explore or go directly to specific data. It will automatically feed into Research Data Australia, the data portal established by



An example of the Data Access Portal search data interface

ANDS, as well as relevant discipline based portals. Entries link directly to download facilities. While some data will be restricted, most data sets will be available to external users. A key feature of CSIRO's system is the focus on ensuring the appropriate level of access to data through access and approval workflows. Where data is restricted to collaborators located in other Australian research organisations, access controls will be enabled using the Australian Access Federation.

"The software aims to ensure data quality and that all the appropriate safeguards are in place for protection of intellectual property and access. There are a series of steps, which require the approval of an authorised person, who verifies that that data is appropriately available," says Miller.

At present, a series of pilot projects and early adopters, including the pulsar data, can be accessed through the portal, and the new enterprise solution will be available from November. CSIRO will be demonstrating the system at the upcoming eResearch Australasia Conference.

"CSIRO's progress in research data management has been significantly enhanced through its relationship with ANDS and the Australian research data management community generally," says David Toll. "The launch of the enterprise level Research Data Service will be a major milestone for CSIRO in growing its research data management capability and we look forward to building the service in the future."

Observing ANZAC Day and documenting cultural heritage - research data management in action

Tim Thwaites, Science in Public

Researchers documenting how we observe ANZAC Day are discovering the value of planning the organisation and sharing of their data, early in the life of their international project. The Faculty of Arts at Monash University has received a large grant to look at how ANZAC Day is observed and reflected on, throughout Australia. It is a long-term project involving many people around the world, and Monash University Library has been running workshops in data management for them.

"Organising data properly at the beginning of a project is absolutely a pre-condition of its effective use, discovery and reuse later," says Cathrine Harboe-Ree, the University Librarian. In this case, for instance, the researchers will be conducting interviews and will need to get appropriate permission if the material they collect is to be used by others or made available for future research.

"Organising knowledge is a traditional role of libraries," Harboe-Ree says. In the electronic age, this role has now extended into the area of data management. "We are interested in getting as much of that data as possible exposed and discovered."

So Monash University Library has moved into advising researchers and research groups on data management or rather, as Harboe-Ree puts it, data *sharing*. "It helps to avoid barriers to the effective use of data." In addition to the ANZAC project, for example, library staff have been working with the whole of the Pharmacy Faculty, providing training in data management for research students and staff, and supporting their strategies to manage data more effectively.

The library is also the custodian of the University's digital archive or repository, which already includes 53,000 items including research data, the University's patent collection, archives of journals for which the University is responsible, PhD theses, research collections and background information on all of them.

Metadata from this repository is regularly harvested by Research Data Australia: the ANDS portal, making it possible to discover some amazing resources. A recent ANDS-funded project has developed software to capture the descriptive data accompanying a set of 7,500 images by photographer John Gollings of the Kashgar region of China, which is ethnically Uighur. The availability of this data will be of interest to researchers of other topics such as architecture, art history, history, cultural heritage, anthropology, sociology, economics and policy. Significant benefits would also accrue to the wider community, including the people of the Kashgar region. The Monash Asia Institute has undertaken to work with

local museums, curators and keepers of cultural relics in Xinjiang to ensure that their work receives international recognition and the public availability of the data would support this objective.



Photograph from the Monash Asia Institute and Urumqi Normal University's research collaboration to document, measure, and define significant cultural monuments and spaces of Kashgar. © John Gollings, 2008. Reproduced with permission

In pursuing its data management role, the Monash University Library works closely with two other units—eSolutions, responsible for providing data storage; and the eResearch Centre, which provides advice on the effective use of technology for research. Working through partnership and collaboration with other university service units as well as the researchers themselves, is also very much part of the approach at the University of New South Wales (UNSW), says University Librarian, Andrew Wells, who believes that management of research data is something, which should "sprout organically, and not be imposed".

But he is much less comfortable when it comes to primary research data. "I am very cautious about collecting and storing such information. The primary research data world is a different one to libraries, which tend to make stuff available to everyone in perpetuity. There are a whole lot of stewardship issues and technical requirements. Some data may only be available to limited numbers of people, and not all data needs to be kept forever."

His views of the important role modern university libraries can play in collecting and managing information—metadata—about the institution's research output reflect those of his Monash colleague. UNSW has benefited greatly, he says, from participating in an ANDS-funded project, which allowed it to generate software to capture metadata about the University's research collection in a form, which could be harvested by Research Data Australia.

ANDS out and about - report on completed ANDS-funded projects

Accessing Architecture Collections

Andrew Williams, ANDS

In August, the Architecture Museum at the University of South Australia successfully completed an ANDS-funded project to improve the management of Museum collection records and significantly increase their visibility online.



Dr Julie Collins, Collections Manager at the University of South Australia's Architecture Museum, with some of the Museum's holdings

The Architecture Museum's data collections differ from many collections around which ANDS projects are based; they are all hardcopy artifacts, including drawings, models, and prints relating to private architectural practitioners. This material first began to be collected in the mid 1970s by architectural historian, Donald Leslie Johnson, and the collection has now grown to over 200,000 items.

Collections and data held by the Museum primarily attract interest from architects, planners, builders, humanities scholars, historians, students, archaeologists, heritage professionals and genealogists. They generally seek information on either an individual architect or firm, or a particular building by an architect.

Museum records were previously online in the form of PDF finding aids on the University of South Australia website. Architecture Museum collections are now visible through Research Data Australia, and also through an online interface to Metatecture, the system developed during the project. The main benefit of the harvest of metadata about the Architecture Museum's holdings into Research Data Australia is increased discoverability. When a Google search takes people to the Research Data Australia page of one of the Museum's collections, they

can use the links to find out more about the Architecture Museum and its other collections, as well as link to further information about the architect whose data they are looking for. It is a much more interactive process now, enabling researchers to make more links, both within the data and to other data about the research topic.

The Museum has already digitised some artifacts in response to user requests, and it is anticipated that demand for digitisation will grow now that the collections are visible online, and can be discovered by researchers who cannot physically visit the Museum in person to see artifacts.



Some items from the University of South Australia's Architecture Museum's data collections

Architecture Museum staff sees this successful project as the start of much more online work around architecture, involving researchers and other institutions with holdings related to architecture. They see the potential for a single portal view onto all relevant holdings, possibly based on a roll out of Metatecture into other organisations. This would amount to a virtual Australian museum of architecture, and such a portal might also act as a locus for collaboration around architectural research through facilitating and hosting various research-related activities. The research would not necessarily be restricted to architecture; architectural data is also of value to researchers in other disciplines, including history, other design disciplines, and research related to any activities conducted in and around the built environment.

Metatecture: <http://www.metatecture.unisa.edu.au/Public/>

Architecture Museum: <http://www.unisa.edu.au/artarchitecturedesign/architecturemuseum/default.asp>

Architecture Museum on ResearchData Australia: <http://services.ands.org.au/home/orca/rda/view.php?key=102.100.100%2F4484>

New Data System Impacts National Climate Research

Debra Truin, Monash University

Many accolades may come to highly successful researchers, but the core motivation of good research is overwhelmingly a passion and dedication for the particular disciplines in which they participate.

Monash University climate scientists, Associate Professor Jason Beringer and Dr Peter Isaac from the School of Geography and

Environmental Science, were among those who answered the call from the ANDS in 2010 to identify key software development projects aimed at solving data management problems for major projects.

A total of eight data capture projects were funded by ANDS at Monash University involving partnerships with the Monash e-Research Centre (MeRC) and the Monash University Library. Of these, \$150,000 was allocated for the capture and publication of Australian ecosystem data through the OzFlux network of measurement sites.

"Initially our motivation was to assist Monash researchers involved in OzFlux..." Dr Isaac said.

OzFlux (<http://www.ozflux.org.au/>) is a \$2.3m initiative of the Terrestrial Ecosystem Research Network (TERN) under the Commonwealth's NCRIS funding initiative of 2008. It aims to increase the number of ecosystem measurement towers throughout Australia to up to 20 sites, enabling broader national climate measurement, and research into energy, carbon dioxide and water exchanges between the biosphere and the immediate atmosphere.

"...however, to ensure the OzFlux project would succeed in the longer term, clearly demonstrating a return on tax-payer investment, there was a need to create a coordinated data system; capable of downloading, archiving and appropriately publishing the data online," Dr Isaac said.

"Part of the ANDS-MeRC solution was to ensure data could be securely archived and backed up at the Monash Large Research Data Store infrastructure (LaRDS)."

Training time for new, less-experienced site scientists who must pre-process the data has been dramatically reduced by standardised online and open source solutions, and the data format, Network Common Data Form (NetCDF), is also commonly used by climate researchers.

Of the recent TERN-funded projects, OzFlux will be one of the first with the ability to automatically register data with Research Data Australia, ensuring it is discoverable by broader national and international research communities.

"Improving the accessibility of OzFlux information is certainly exciting as we are part of the larger international FluxNet network, while improved accessibility to climate data is critical if we as a species are to have an opportunity to understand the enormous impact we are having upon our planet," Dr Isaac said.

Chair's report – Ron Sandland

The importance of the ability to access and use large and complex data was at the fore in a conference Ross Wilkinson and I attended recently in Adelaide. The conference was the National Data Linkage Forum sponsored by the South Australian Department of Education and Children's Services. Its theme was ambitious: using administrative data to improve education outcomes for Australians. Many ideas emerged from the conference but none more important than using data to drive policy formation. Global studies clearly link critical national outcomes: health; prosperity; social harmony, and productivity to early (indeed very early) educational experiences of the developing child. These decisions will underpin the maintenance of civil society. But gathering, collecting, accessing, linking the data that exists requires a data infrastructure to enable it to happen.

The data resources that already exist (e.g. the National Assessment Program–Literacy and Numeracy (NAPLAN) and the Australian Early Development Index (AEDI)) will go a long way to providing the basic resources needed to do this. But like health data, which will itself form an important part of the data estate, confidentiality requirements will dictate the need for a careful approach and privacy-preserving data linkage techniques will form an important part of the toolkit.

There are many hurdles to be overcome before this data-driven world can take shape, such as the development of priority tasks and data analytic skills in the educational research community. But previous Department of Innovation, Industry, Science and Research (DIISR) investments, through National Collaborative Research Infrastructure Strategy (NCRIS) and Super Science—Population Health Research Network (PHRN), Australian National Data Service

(ANDS), National eResearch Collaboration Tools and Resources (NeCTAR) and Research Data Storage Infrastructure (RDSI)—means there is already significant infrastructure capability to build on. And the warm support, in their respective opening and closing speeches, by the South Australian Premier–Elect, Jay Weatherill and Federal Minister, Kate Ellis, were very positive indicators.

The exciting world of data possibilities was given a new twist at the forum in the terrific talk by Dr Nick Gruen. He described Kaggle (<http://www.kaggle.com/>) which "is a platform for data prediction competitions that allows organisations to post their data and have it scrutinised by the world's best data scientists". It gives a really interesting view into a world where data and the predictive aims of a sponsoring organisation are posted with a prize, in one case \$3 million, for the best solution. In some ways the idea is not unlike that of Linux, where many software developers contribute in a competitive way to the Linux core. Many of the best solutions to have emerged in Kaggle have been from people with no formal data analytic training.

But none of this is possible without the ability to provide access to the data. And that is where capabilities like ANDS come in.

In my last Share article, I said:

"With appropriate investment in national infrastructure, the data-intensive science of the future will be an exciting place for researchers."

I think those of us who work with ANDS can be rightly proud of the role we can play in making this world a possibility.

ANDS event reports

Digital Preservation Management Workshop

The ANDS sponsored the highly successful Digital Preservation Management Workshop, in Melbourne on 20 and 21 July 2011, which was facilitated by Dr. Nancy Y. McGovern, Digital Preservation Officer for The Inter-university Consortium for Political and Social Research (ICPSR), and Kari R. Smith, Digital Preservation Management Workshops Project Manager at ICPSR. The 26 participants were drawn from universities, research organisations and government departments.

The workshop focused on preservation of digital objects, whether they were 'born digital' or not, and the two days followed a well-developed sequence, which was designed to assist institutions in establishing a policy framework. The framework was populated with the requisite organisational and technological components in parallel, with each informing the other.

Participants very much appreciated the wealth of documentation, which accompanied the workshop—both the pre-workshop tutorial and the incredibly comprehensive manual. Openly available support materials included:

1. The Digital Preservation Management tutorial (http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html) provided background and introductory information on digital preservation management issues, including terms, concepts, and an overview of the digital preservation model that is at the core of the workshop program.
2. Trusted Digital Repositories: Attributes and Responsibilities (<http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>)
3. Open Archival Information System (OAIS) Reference Model (<http://public.ccsds.org/publications/archive/650x0b1.pdf>) or the more concise overview: "The Open Archival Information System Reference Model: Introductory Guide" by Brian Lavoie at: http://www.dpconline.org/docs/lavoie_OAIS.pdf

All of the participants rated the overall workshop highly, and noted that the most useful components were:

- » developing a strong understanding of the OAIS (Open Archival Information System) model
- » working through preservation metadata management issues
- » the resources available to support the policies and resourcing associated with data preservation governance

ANDS has a number of resources on preservation, beginning with Creating a Data Management Framework (<http://ands.org.au/guides/dmframework/data-management-framework.html>)



L to R: Kari R. Smith, Karen Visser and Dr. Nancy Y. McGovern at the ANDS Digital Preservation Management Workshop

"The workshop focused on preservation of digital objects, whether they were 'born digital' or not..."

Conference reports

eConcertation event and All Hands Meeting 2011



Cité internationale, Lyon
Copyright UMR CNRS MAP - ENSAL Centre de Congrès de Lyon

In late September, Andrew Treloar, Director, Technology at ANDS, attended an eConcertation event and the All Hands Meeting in Europe.

The eConcertation event was held in Lyon, France, and had two sessions of particular relevance to ANDS: Scientific Data and e-Infrastructures future prospects under HORIZON2020 (the new name for what was called Framework Program 8).

The Scientific Data session highlighted a series of European Union (EU) reports on data challenges over the last four years, including the Riding the Wave report, which was submitted to the European Commission in 2010. A key theme in this report was the idea that “Data themselves become the infrastructure”, however making this possible means meeting the challenge of fostering interoperability between e-infrastructures for data across institutions, disciplines and countries—while keeping the balance between vertical discipline solutions and horizontal enabling layers. The current Framework Program 7 is funding two attempts at meeting this challenge. The first is the OpenAire+ project, building on an existing network of institutional document repositories to add mechanisms for dealing with data. The other is a project called EUDAT which is seeking to identify and offer generic data services by extracting them from a number of existing integrated discipline solutions for data management. Of note in the session was the foreshadowing of a possible pilot in HORIZON2020 that will require managing data outputs and making them available as a condition of funding. This approach is consistent with the ways that Open Access to the outputs of research is now increasingly anchored in EU policy. The Scientific Data session also contained a panel discussion, to which ANDS was invited to contribute, looking at the views of the Data Interoperability Task Force on data infrastructures.



Ron Cooke Hub, University of York
CC-BY (<http://www.flickr.com/photos/robbphotos/>)

The session on e-Infrastructures made the point that e-infrastructures needed to balance a number of tensions: flexible yet reliable, secure yet open, local and global, and affordable and high-performance. The bulk of this session dealt with the mechanics of applying for a number of calls that the EU has open, dealing with international cooperation with the USA on data issues. The EU was keen to have an Australian perspective on these calls, which is why an ANDS representative was invited to speak.

The All Hands Meeting (<http://www.allhands.org.uk/>) was held at the University of York in the city of York, United Kingdom. Previously, this annual gathering was for the UK e-Science program, which is now finished, therefore the focus of future All Hands Meetings is changing, starting with next year’s theme—Digital Research. A number of presentations provided useful pointers to possible projects that could contribute to software for ANDS-funded Applications activity. In addition, there were opportunities for ANDS to renew connections with key UK contacts and international collaborators. The first meeting, with Simon Hodson who manages the JISC Managing Research Data program, was an excellent opportunity to review current data priorities in the UK and Australia, and to plan future collaboration. The second meeting, with Kevin Ashley who directs the Digital Curation Centre, allowed for half a day working through the progress on actions arising from the ANDS sponsored invitational workshop held in April in Prato, Italy (sometimes called the #dataprato event, after the Twitter hashtag).

Both meetings demonstrated once again the high regard in which Australia’s data management activities are held, and our leadership in national data infrastructure.

DataCite Summer 2011 Meeting and CODATA Data Citation Workshop

Adrian Burton, Director, Services at ANDS, recently attended two meetings in Berkley, California addressing the issue of data citation.

The first was a small invitational workshop convened by the USA National Academies, the CODATA committee of the International Council of Scientific Unions, and the International Council for Scientific and Technical Information. These bodies are seeking to develop very high-level standards and practices on data citation. As such the workshop sought to identify fundamental scientific drivers for citation such as attribution and reproducibility. At the same time the group also addressed obstacles or discipline-specific requirements. The deliberations of the workshop will inform broad policy statements by the convening bodies.

There is a growing desire in the research community to link data and publications as well as to measure and acknowledge the impact of research data—hence the interest in data citation from these

top-level science funding and policy bodies. Reports are expected over the next 12 months.

The second meeting was the Summer Meeting of the DataCite consortium, which promotes the citation of research data in similar ways to which journal articles are cited. ANDS belongs to this consortium, which enables it to supply Australian data centres with Digital Object Identifiers (DOIs) to optimise the automated tracking of citations in the world of scholarly communications.

The DataCite meeting had a much more hands-on approach with presentations by data centres and libraries active in data citation. The US library system requires librarians to publish for tenure and promotion; there were a number of interesting presentations on studies of citation behaviour, journal policies, institutional support, and so on.

Meet the ANDS Staff

Karen Visser



Karen Visser is Program Leader, Skills, Resources and Policy for the Capabilities and Frameworks Program. Karen is based in the ANDS Canberra office. Capabilities activities focus on building the capability of researchers and support staff (including ANDS staff) to contribute to, and better exploit, national data infrastructure. The Program also works with research communities and local e-Research support services to improve particular

data-related competencies, as well as enhancing and adding national focus to institutionally based support, materials development and training initiatives.

Karen comes from a teaching, teacher education, training, library, project management and information management background. Immediately prior to joining ANDS Karen spent six months on a

remote atoll supporting teacher education in the Maldives. One of the benefits of this varied background means that Karen has a deep understanding of the myriad of ways in which change is facilitated.

The Capabilities workplan Karen has devised for the next six months is full and varied. One highlight is the creation of a number of Information Packages to support the various ANDS Services, starting with DOIs (Digital Object Identifiers) for Data Citation. The concept she has envisioned is simple, but if it is as successful as hoped, will help ANDS services to be more readily adopted into the research community. Simply put, all the information an individual or organisation needs to put in place a particular ANDS service will be accessed from one page on the ANDS website. Information for each ANDS Service will include the following: What is it? What's in it for us? How does it work? What do we need to do to make it happen? What support, documentation and training can ANDS offer? Who else is using it and what is their experience? Is it in project or production mode? What are other related things that might be needed (policies, frameworks, allied technologies etc)? This is an exciting concept and augments the numerous other ways in which ANDS engages with our partners.

In brief

A Starring Role!

An ANDS-funded project was a recent winner at the prestigious CSIRO Awards 2011. The Parkes Observatory Pulsar Data Archive Team won the CSIRO Medal for Support Excellence for the development of a data archive that captures data directly from the CSIRO Parkes radio telescope, stores it and feeds metadata into the CSIRO repository, discipline portals and into Research Data Australia. The metadata can be viewed and the data accessed via Research Data Australia (<http://services.ands.org.au>) and the CSIRO Data Access Portal (<http://datanet.csiro.au/dap/>).

The team was a partnership between research scientists, librarians and IT staff in CSIRO and the combination of skill sets was a major factor in the success of the project. More information on the award can be viewed on the CSIRO web site (<http://www.csiro.au/science/2011-Support-Excellence-Medal.html>).

The Australian National Data Service software update - release 6.1

The implementation of Release 6.1 of ANDS software is now complete. All services, including a new look Research Data Australia (<http://services.ands.org.au>), the ANDS Collections Registry and Identify My Data (<http://ands.org.au/guides/identify-my-data-awareness.html>) are now online and available. This release features:

- » a completely new version of Research Data Australia with new page layouts and improved search and browsing capabilities
- » a new alternative to the Sandbox registry system with increased support for managing, assessing and publishing draft records

DIISR 2011 strategic roadmap for Australian research infrastructure

On 29 September, 2011, the Department of Innovation, Industry, Science and Research released the 2011 strategic roadmap for Australian research infrastructure. The roadmap has been developed through extensive consultation with researchers and other stakeholders, and builds on the roadmaps from 2006 and 2008. It reinforces the themes of innovation driving competitiveness and productivity, and highlights the effectiveness of a collaborative

approach to determining the key pillars of research infrastructure. The consultation process revealed the growing and insistent need for integration to connect current and future facilities to support the provision of seamless generic and specific research infrastructure services.

<http://www.innovation.gov.au/Science/Documents/2011StrategicRoadmapforAustralianResearchInfrastructure.pdf>

"Research Data Australia is not only a discovery portal, but also a publicity portal. It's a way to make research data visible to others."

Dr Ross Wilkinson

Forthcoming events

ANDS activities at eResearch Australasia, 6-10 November 2011

The Australian National Data Service will be hosting a booth at eResearch Australasia, where delegates can interact with ANDS staff and help us build a 3-D representation of the Australian Research Data Commons using Lego bricks as building blocks. ANDS will also be involved in the following activities:

Monday 7 November, 2011: Conference Day 1 from 2:55pm to 3:55pm
Dark Data and the Long Tail of Science—Birds of a Feather session
Lesley Wyborn, Bryan Heidorn and Andrew Treloar
More details: <http://conference.eresearch.edu.au/presentations/#99>

Monday 7 November, 2011: Conference Day 1 from 5:00pm to 6:15pm
Persistent Interconnection of Publications, Data and Computational Models through Standard Citations—Poster session
Gad Abraham, Jeff Christiansen, Ross Wilkinson and Justin Zobel

Tuesday 8 November, 2011: Conference Day 2 from 3:25pm to 4:25pm
ANDS Projects—Birds of a Feather session
David Groenewegen
More details: <http://conference.eresearch.edu.au/presentations/#6>

Tuesday 8 November, 2011: Conference Day 2 from 8:30am to 8:50am
Licensing Australia's Research Outputs Under a Single Framework—Presentation
Baden Appleyard (AUSGoal)
More details: <http://conference.eresearch.edu.au/presentations/#45>

Wednesday 9 November, 2011: Conference Day 3 from 12:00pm to 12:20pm
Costs and benefits of data provision—Paper presentation
John Houghton and Greg Laughlin
More details: <http://conference.eresearch.edu.au/presentations/#16>

Thursday 10 November, 2011: Day 4 from 9:00am to 12:30pm
Building a culture of research data citation—Half Day workshop
Jan Brase, Stuart Hungerford, Karen Visser
More details: <http://conference.eresearch.edu.au/workshops#8>

Friday 11 November, 2011: Day 5 from 9:00am to 4:00pm
Scholarly Infrastructure Technical Summit—Full Day workshop (Invitation only)
Andrew Treloar

More details regarding the program: <http://conference.eresearch.edu.au/program-2/>

Further information about the eResearch Australasia conference is available at: <http://conference.eresearch.edu.au/>

Get notified about our forthcoming newsletters via RSS feed: <http://ands.org.au/newsletter>. For more news, alerts, announcements and discussion subscribe to the ANDS General group: <http://groups.google.com.au/group/ands-general/subscribe>



ANDS Office
680 Blackburn Road,
Clayton VIC 3168
Telephone: 03 9902 0585
Email: contact@ands.org.au

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy program and the Super Science Initiative.
This newsletter is designed by AK Design (<http://www.akdesign.com.au>)

ANDS Project Partners:



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 2.5 Australia License <http://creativecommons.org/licenses/by-nc/2.5/au/>