

Piecing together the eResearch puzzle

Ross Wilkinson, ANDS & Tim Thwaites, Science in Public



Image courtesy of Paul Trafford (CC BY 2.0)

Imagine having the best analytic tools processing the most complete data sets running on one of the fastest computers in the world in collaboration with colleagues connected by high-speed networks into a virtual laboratory. This is not a dream for an increasing number of Australian researchers. The alphabet soup of facilities that comprise the nation's investment in research infrastructure—ANDS, NeCTAR, NCI, NRN, RDSI, BPA, TERN, IMOS and many others—is already delivering it.

Institutions get significant value out of this set of facilities. The University of Tasmania, for instance, plays a leading role in marine research. That capability is founded on access to the data being collected by Australia's automated Integrated Marine Observing System (IMOS) and other institutions across the country. A virtual laboratory in the areas of marine and ocean-climate science is now being established with funding from the National eResearch Collaboration Tools and Resources (NeCTAR). This will make it possible for researchers to use an intuitive web application to configure a range of ocean and wave models efficiently, with specifications of choice, for any region and time period. It's an environment that allows marine researchers at the University of Tasmania to engage with the best in the world.

"This on-demand modelling of ocean currents driven by comparisons of the simulations with real ocean observations means we can test hypotheses, discover new knowledge, and make decisions more rapidly than ever before," says Prof Nathan Bindoff, Research Program Leader—Climate Change and Ocean Processes at the University of Tasmania's Institute for Marine and Antarctic Studies (www.imas.utas.edu.au).

Researchers often need more than their own data and more than their own computer. For example, even to determine local conditions, climate scientists work on models that have thousands of data streams integrated from around the world. The datasets they now use are measured in petabytes (millions of gigabytes), and that size stresses our capacity to store, process and transmit them. So the focus of the National Computational Infrastructure (NCI) National Facility (at ANU) on tools, computation and storage of climate data provides a unique advantage to Australia's climate scientists.

In fact, the efficient storage of large sets of data to allow future access is a whole specialist enterprise in itself. "This data needs to be stored somewhere if it is to be available for reuse and collaboration by researchers. That is the role of Research Data

Inside Issue 14 — Data as a strategic asset

- » The useful whole is greater than the sum of the data
- » We all benefit from free public information
- » Institution profiles
- » ANDS events reports
- » Modelling the movement of malaria
- » Managing research data outputs
- » Chair's report
- » In brief
- » International engagements
- » Forthcoming events
- » ANDS-funded projects

Storage Infrastructure (RDSI)," says Dr Nick Tate, Director, RDSI Project. "RDSI is working with a number of partners across the country to develop a massive shared storage capability."

Blogging after three days at the 2012 International Conference on Research Infrastructure (ICRI), the director of the Australian Research Council's Centre of Excellence for Climate System Science, Prof Andy Pitman observed:

"What came as a surprise to me at the ICRI conference is that, while a country might have a well-defined program around one of the components of eResearch, no country appears to have all the

pieces of eResearch under control. So, while a country might be better than Australia at one component of eResearch, no country has established and linked all the pieces of the e-jigsaw well.

In terms of climate science, NCI provides our computer resources and RDSI our storage needs. ANDS helps us understand how to manage our data and publish it. NeCTAR provides tools to help process data, to manage codes and to develop an efficient workflow system (a virtual laboratory if you like). Australia has put in place the pieces of the jigsaw in a way that really is world-leading and unique in covering all the key components."

"...Australia has put in place the pieces of the jigsaw in a way that really is world-leading and unique in covering all the key components."

The useful whole is greater than the sum of the data

Ross Wilkinson, ANDS & Tim Thwaites, Science in Public

We now possess a vastly improved capacity for handling information, and so the research data assembled by an institution can be a key asset. Used effectively, it can play a strategic role in informing its research planning and output. Such a strategy might focus around areas of research strength, the ability to collaborate and form research partnerships, or issues of particular relevance to the institution. Sometimes all of those strands can be supported simultaneously.

The tropics are a major focus of James Cook University. Thus, it uses its Tropical Data Hub as an asset—to enhance its research capability, to develop relationships with universities that have a similar focus and to support research into significant problems such as tropical diseases or the consequences of climate change on the tropics. By creating and enhancing a data collection of significance, a research institution can build a natural advantage, showcasing its capability as well as an institutional asset.

Sometimes data collections are already of such a scale and complexity that they constitute a national or even international asset. At other times, an institutional collection needs the critical mass of a national context to enhance its value. It can then act as a seed to attract data from sources outside the institution.

For instance, Australia's expertise in climate change adaptation research is multi-faceted, multi-disciplinary and multi-institutional. Griffith University is working with ANDS to connect collections of climate adaptation research data, such as health impacts of extreme temperatures and mammal refugia. While the individual collections are of great value, the linked collections are growing into an asset of

national significance and providing Griffith with a resource on which to base its own climate change adaptation research. The value goes beyond that. Much of the crucial data for climate change adaptation research is held by government departments, local agencies, and private industry. Connecting all this data makes it much more useful. Connecting to the big climate models and their outputs generated at the National Computational Infrastructure (NCI) makes it more useful (see p.5). And all of that leads to new uses of this data.

Such a research data asset not only gives an institution's researchers a unique tool that provides them with a distinct advantage, it also enables them to collaborate with the best around the world.

This is where connectivity, publication, discovery, and reuse of the data become crucial, and where the specialised expertise and services of a facility such as ANDS can add great value. By utilising Australia's other eResearch infrastructure capabilities, it is now possible to integrate and analyse data with new tools and powerful computation, to store the results, and then to integrate them into an even more powerful collection.

The national aggregation of these institutional assets, all forming part of a collaborative research data commons, provides an advantage to the Australian research system as a whole—enabling it to tackle problems on a national and international scale, to assemble data in ways that allow many different disciplines to contribute, and to ensure that Australian researchers become international partners of choice when embarking on data-centric research.

Modelling the movement of malaria

Tim Thwaites, Science in Public

Within a year, researchers from all over the world should be able to visit an Australian website to model or test control strategies for the transmission of malaria in a specified area at any time of year—at least that's what Prof Tom Burkot of James Cook University (JCU) in Cairns hopes. He's the Orchestrator of the Vector Ecology and Control Network (VECNNet)—a web-based project—which is putting JCU at the very centre of international efforts to control and eliminate malaria.

The VECNNet strategy is to: bring together a library of data and information—on malaria and the habits and ecology of its carriers, the hundreds of species of Anopheles mosquitoes that transmit it; organise, store and make this data accessible; and provide a test-bed for intervention—with tools and predictive models so the data can be analysed and suggest answers to questions as to the best and most economic ways to control a disease that kills more than 650,000 people worldwide each year.

The project, which initially sprang from the Bill and Melinda Gates Foundation, now has partners that include the universities of Pittsburgh and Notre Dame in the US, Oxford in the UK, the Swiss Tropical and Public Health Institute, R. Farlow Consulting from Texas and the Gates-inspired Intellectual Ventures Laboratory.

A mixture of serendipity and 'making your own luck' brought VECNNet (www.vecnet.org) to JCU. At the time the founding partners decided to establish VECNNet, they also agreed on a leader—Tom Burkot, an expert on the spread of malaria based at the Centers for Disease Control in Atlanta. When he told them he had just accepted a position with the Queensland Tropical Health Alliance at JCU, the project came with him.

And it fell on fertile ground. The University has a tropical focus, and Cairns is one of the few places in Australia with several people who were respected in the field of vector biology. "There was a natural hole for VECNNet to fill," Burkot says. In addition, and as equally important, JCU was a centre of information technology. It had its own eResearch centre and could also count on the support of national facilities, such as ANDS. "The support for computing and storage through these facilities is really fantastic," Burkot says. "I think it will allow VECNNet to be more easily sustained in the future."

"One of the real benefits [to JCU] of this very high profile project—which really has captured the imagination of people—is the extent of collaboration. We probably have hired about 15 people full-time, but there are another 50 individuals at our partner institutions spending significant amounts of time on VECNNet. And one of the founding principles is that we want to be transparent and share information. Anyone who is a partner of VECNNet has access to all of the intellectual property that we develop."



Image of an adult mosquito (male) courtesy of Spike Walker, Wellcome Images (CC BY 2.0)

VECNNet is one of several projects or collections managed by Australian institutions that involve a high level of international interest. They include:

- » The Human Chromosome 7 Proteomics project at Monash University
- » The National Linguistics Corpus at Griffith University
- » The European Molecular Biology Laboratory mirror site at the University of Queensland
- » The Census Data of the Australian Bureau Statistics
- » The Repository of Antibiotic Resistance Cassettes at the University of New South Wales
- » The Parkes Observatory Pulsar Data Archive at CSIRO
- » The Massive Mouse Brain Map at the University of Queensland
- » The Life Patterns project at the University of Melbourne
- » The Integrated Marine Observing System datasets
- » The Geoscience Australia collections
- » The SkyMapper project at ANU
- » The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) at the University of Sydney.

We all benefit from free public information

Tim Thwaites, Science in Public

After much debate, the Australian Government agreed in 2005 to fund the Australian Bureau of Statistics (ABS) to make its information available online for download free of charge. In 2008, the ABS introduced standardised Creative Commons licensing, allowing unfettered use of its data. A cost-benefit analysis of the provision of public sector data (ands.org.au/resource/cost-benefit.html) by Prof John Houghton from the Centre for Strategic Economic Studies at Victoria University showed that, while the ABS lost revenue when it stopped selling data, it also made substantial savings because it no longer needed a 'shopfront' or to spend time on licensing negotiations or related enquiries. In fact, Prof Houghton found that the benefits of making ABS data freely available and with an open licence were likely to have been more than five times the costs, with the nation starting to put the information to wider use.

Census information, in particular, is prized. It provides a snapshot of many facets of the Australian population—age, sex, location, wealth, education—that companies, government and organisations can put to profitable use. Throughout the most recent five-year census cycle—from the 2006 census to the 2011 census—downloads have not dipped below about 1.5 million a year, three times the average figure before free availability. But when the census data was released in 2007-08, there were more than three million downloads.

In another analysis, Prof Houghton estimates the free provision of spatial data by Geoscience Australia since 2001 has been even more valuable to the nation.

While it is difficult to assess the benefits of making public sector data freely available, Prof Houghton developed an approach that provides strong policy guidance. "Much public sector information, such as that from ABS and Geoscience Australia, is collected in the course of government business," he says. "The cost of collection is what economists call a sunk cost, it has already been spent and cannot be recovered. By simply making the data freely available online and open to reuse, the government and nation benefit through increased economic activity and taxation revenues."

Noting Prof Houghton's argument, the Victorian Government has recently announced policies to allow 'unprecedented access to the state's data' online.

Cost and benefit of data provision		
Benefit	=	Agency savings + User savings + Increased returns to annual expenditure on PSI production
Cost		Agency costs + User costs

Managing research data outputs & inputs

Cynthia Love, ANDS

Have you ever negotiated access to a dataset only to discover that one of your colleagues has jumped through all the hoops already? While there is considerable and justified focus on the publication, management and reuse of data that are outputs of research, there is also a strong case for examining the management of data that are inputs to research.

There is any number of stories about researchers in the same institution acquiring duplicate copies of datasets. In certain circumstances this makes sense, as the data needs to be near the tools and compute. In the changing environment of cloud computing there will be less need for this. Accidentally duplicated acquisitions are a more likely scenario. An argument for greater economic efficiency certainly can be made but in an environment of increasing open access to data there are other good reasons to manage the data inputs into research.

If organisations are increasingly viewing data as an asset, management of the licences of the sets that have been acquired also becomes important. It does not necessarily make sense for an organisation to maintain multiple licences for the same dataset. Maintenance of the data holdings and associated licences and making these

datasets visible to users can be done through institutional metadata stores such as is the practice in CSIRO, like the Data Access Portal (data.csiro.au).

Another administrative driver for improved practice would be efficiency in storage and retrieval of data especially where large datasets are concerned. As the data deluge continues it is important to manage storage efficiently to enable cost-effective storage and retrieval.

Improving the administration of acquired/licensed data holdings sets the foundation for improved discovery of what is held in an organisation and can lead to unexpected collaboration and knowledge sharing. However one of the most critical reasons for managing the data inputs to research is the management of provenance for derived data and citation. This is a significant factor in research integrity, particularly where the provenance of a derived dataset becomes complex with a number of different licensing regimes for source data. Equally important is the citation of other input data; just as one manages the references to publications in the literature, acknowledgement of the data inputs should also be managed. As an input to research, and as one of its outputs, data needs to be managed and managed well.

Institution profiles

Bringing together big data at ANU

Ross Wilkinson, ANDS

What happens when researchers have access to the computational resources, the tools, and the data to tackle really big problems? There is a nexus of resources in Canberra that will enable researchers from around the country to tackle some of the biggest problems in research.

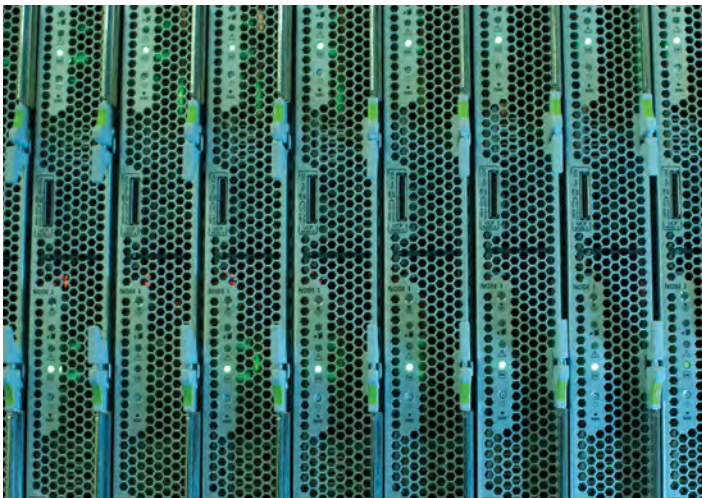


Image of the NCI supercomputer, courtesy of NCI

Prof Lindsay Botten, head of the National Computational Infrastructure (NCI) at ANU notes that the most detailed map of the heavens ever compiled, charting a vast dome of stars extending from the equator to the South Pole, is being created with the help of the NCI. Nobel Laureate Prof Brian Schmidt describes harnessing a new telescope with processing power: "We are using the new fully robotic SkyMapper telescope—the widest-field instrument in the world of this size and producing torrents of data, 225 terabytes in all—which is why we need the phenomenal processing power of the NCI supercomputer". The addition of the NeCTAR-funded All-Sky Virtual Observatory means that both the processing power and the tools to exploit the internationally significant dataset will be available, providing a great research advantage to Australian researchers and their collaborators.

When the world's top climatologists gather in 2013 to report on how the earth is changing, predictions made using the most powerful climate model ever built in Australia will provide vital Southern Hemisphere input to the global picture. The models running on the NCI supercomputer, together with new tools being developed at ANU in the NeCTAR-funded Climate and Weather Science Laboratory, as well as crucial data used by the Intergovernmental Panel on Climate Change, provide an incredibly powerful research environment to tackle the world's biggest problems.

Preparing for a data-centric research future

Martin Borchert, Queensland University of Technology

eResearch, the use of computers in research—methods and infrastructure—is also about adopting data-centric methods of research, both analysis and visualisation. Research data are now recognised as valued outputs from the research process in their own right, and not just an intermediate step in the development of the final research publication. National research policy and research funding agencies are increasingly requiring research data to be properly managed and made available. The acquisition of research data is often the most time consuming and costly phase of research work and so it makes sense that the sharing and reuse of research data can lead to faster and more efficient creation of knowledge.

Queensland University of Technology (QUT) has been working with our friends in the Australian National Data Service (ANDS) for over three years now to investigate and develop processes and infrastructure to prepare the University for the ultimate goal of being able to identify, describe, store, share and reuse data created by QUT research—that is, to treat them as strategic research assets.

A series of projects have been implemented to develop the required pieces to complete the service know-how: how to conduct the data interviews; the motivations and disincentives for sharing and reusing data; how to work with researcher to develop software that meets their needs (and captures their data outputs); and about good practice in the description of datasets using the RIF-CS metadata schema. The projects also provided resources to develop the QUT Research Data Finder repository—a data registry for the University—based on the VIVO software, and linking to a variety of existing sources of truth (internal and external) to bring information together to deliver an efficient process and an authoritative knowledgebase on which to build the data management service.

QUT has complemented the work funded by ANDS with its own work on developing advocacy mechanisms and knowledgeable staff, a data management policy, supporting information and seminars and workshops to support the data management service.

It's a long and often complex path, but it's a necessary one to take in order to prepare the University, step-by-step, for the increasingly data-centric present and future.

"Research data are now recognised as valued outputs... in their own right"

Ensuring our data's longevity

Ian Atkinson, James Cook University

James Cook University has always considered its research data to be an important asset. The University, based as it is in the Australian tropics, has always had a focus and intensity on environmental research and translation of this research into policy and management of the unique environments of northern Australia.

A natural consequence of this focus has been the requirement to store primary research data and where possible ensure its long-term storage and accessibility. The recent launch of the Tropical Data Hub (tropicaldatahub.org) represents the coming together of these ideas to provide an integrated solution for researchers to:

- » store primary data sets
- » generate appropriate metadata to enable data and preservation
- » disseminate and discover research results

Of course, as with any new development there is a range of opinions, suggestions and various levels of adoption and enthusiasm apparent, and the Tropical Data Hub (TDH) project is no different to any other. However, it has been very pleasing that a large and diverse range of researchers across disciplines and career levels have embraced the TDH concept as a way to promote and disseminate the fruits of their research.

We are now bedding the TDH concept down into the internal process of JCU while looking to grow and extend the range of organisations involved.

Investing in infrastructure & people

Anna Shadbolt, University of Melbourne

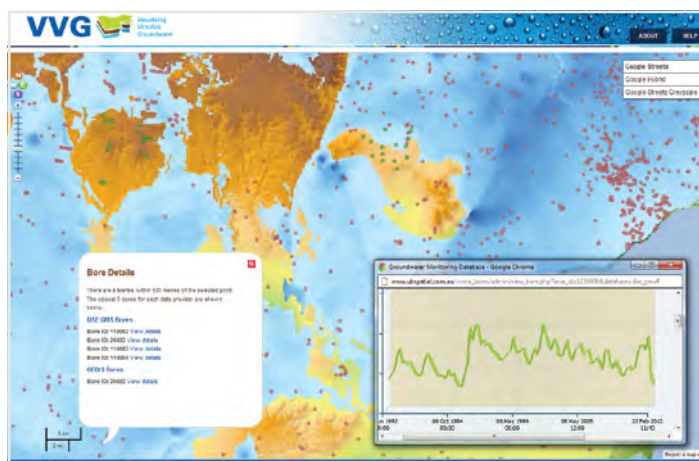
The University of Melbourne started seriously examining research data, information, and records management in 1996, as evidenced in *Hunters and gatherers: from research practice to records practice* (Grady, K., McRostie, D., & Papadopoulos, S. (1997) *Archives and Manuscripts*, Vol 25: No 2, November 1997). Since that time there have been three policy reviews including audits relating to research data management. The University of Melbourne views research data generated, collected and mediated by its researchers as highly valuable and has invested heavily in significant infrastructure that supports good practice in research data and information management, including data storage, central data registry, central research information management system, as well as core teams of expertise across a number of organisational units. Funding from ANDS and other national initiatives has complemented these internal efforts, improving internal practices and accountability.

In 2012, the University of Melbourne has a solid core of distributed expertise dedicated to research support (including research data management). These teams add up to 50+ people across ITS-Research (Central IT), eScholarship Research Centre (Library), Research Data Services (Library) and the Melbourne eResearch Group (Melbourne Research).

Ballarat, digging for data gold

Helen Thompson, University of Ballarat

The University of Ballarat (UB) is increasingly treating research data as a strategic asset with exemplary practices emerging in areas including natural resource management data, regional development, agricultural production, and health and wellbeing.



Screenshot courtesy of the University of Ballarat

A small feasibility study in 2007 provided the initial impetus for establishing UB Spatial (ubspatial.com.au), an interoperable web-GIS portal and groundwater database that was developed with the collaborative support of catchment management authorities, government departments and municipalities, all of whom are end-users of the service. Since that time initiatives have taken place in areas including disaster mitigation, region knowledge management and broadband enabled innovation.

The most significant initiative to date is Visualising Victoria's Groundwater (vvg.org.au)—a research project, which brings together a unique collaboration of regional, state, national and international organisations with the common aim of federating groundwater databases in Victoria.

This \$2.23M collaboration is effectively addressing issues associated with data capture, aggregation, transmission, storage, access, reuse and curation that are identified in the national eResearch infrastructure priorities as essential components in Australia's ongoing research success. In particular, the project provides access to research significant data held outside research institutions by developing the tools for data linking, analysis and visualisation.

A key feature of the VVG system is that data managers maintain custodianship of their data and databases, but allow interoperable exchange by adopting the Open Geospatial Consortium standards for groundwater data exchange—GroundwaterML—developed by Natural Resources Canada.

By providing an effective and easily accessible tool, water authorities are able to better educate their customers on the science of groundwater. There is also potential for broader benefit through the application of the developed technology in virtually federating other environmental, social and economic datasets.

Chair's report—Data as a strategic asset

Ron Sandland, ANDS' Steering Committee Chair

Australian researchers whether in universities or in publicly-funded research agencies are living in an increasingly competitive world. And data has a major role to play in that competition. As a metaphor for this we need look no further than the Shanghai JiaoTong rankings of university academic performance, the assembly of which is a major data collection and interpretation exercise in its own right.

Those who have associations with universities can be in no doubt as to just how seriously these rankings are taken. They affect an institution's ability to attract leading researchers, postgraduate students, post-docs, philanthropic donations, the best and brightest undergraduates, overseas students, international collaborators and industry partnerships, all of which are critical to their ability to thrive. The rankings themselves form part of a positive reinforcement system—without some intervention the rich get richer and the poor get poorer.

The research impact achieved by Australian researchers is known to be significantly higher for those whose publications are co-authored by international collaborators. The ability to effectively capture, analyse, interpret and share data is a fundamental trading chip in generating productive global collaborations. And it is not drawing too long a bow to suggest that this capability will be a significant strategic asset in determining the success of Australian institutions in the global research enterprise.



Australia has been wise in the investments it has made in research data infrastructure encompassing both data management and data generation capabilities which are well recognised internationally. I wrote in the last *share* newsletter of the value to national health policy of the investments being made in the ability to confidentially link and analyse medical data. I could just as easily have written about IMOS, TERN, ALA, AURIN or other data generating capabilities.

The critical nature of research data infrastructure has been recognised in the *2011 Strategic Roadmap for Australian Research Infrastructure* (innovation.gov.au/Science/ResearchInfrastructure/Pages/default.aspx) and the Department of Innovation has recently asked me to chair the Research Data Infrastructure Committee to look at the areas of continuing investment need. The terms of reference of the committee are as follows:

- » Provide a forum for consultation between government, data generation and data management research infrastructure capabilities, and the broader research sector to inform future planning for research data infrastructure in Australia;
- » Develop a framework view of future requirements for research data management in Australia;
- » Advise on the interrelationships between the data infrastructure investments and the relative roles that government, institutions and research facilities can play in optimising these;
- » Advise on the means by which the effectiveness and efficiency of the relationships between data generating investments and the data infrastructure investments could be optimised in future implementation;
- » Propose options for governance models that would optimise the outcomes from future national research data infrastructure investment; and
- » Starting from a position that research data is a strategic resource for both individual institutions and the nation as a whole, we look forward to consulting with the research community to determine how best to address these terms of reference.

Australia has been wise in the investments it has made in research data infrastructure...

International engagement

Research globally, Manage locally

Andrew Treloar, ANDS

Research is an international endeavour. As one example, research into the genome of one of Australia's iconic animals (the Platypus) was carried out by an international consortium consisting of over 100 researchers from the United States, United Kingdom, Germany, Israel, Japan, Spain, New Zealand and (of course!) Australia (genome.wustl.edu/highlights/article/platypus).

So, what role can ANDS play in assisting Australian researchers work in a global research environment?

As the above example demonstrates, researchers need to be easily able to work with international data collections. But they also need to easily be able to work with international research partners over those collections. This is one of the reasons why ANDS is involved in the setting up of the Research Data Alliance (rd-alliance.org), formerly called the Data Web Forum. We expect that the various working groups set up through this organisation will deliver policies, practices, software and tools to enable smoother and richer engagement over research data.

And, of course, ANDS isn't the only group in Australia who cares about this issue. A number of other data-intensive NCRIS

capabilities are also actively collaborating with international colleagues. IMOS (Integrated Marine Observing System) is heavily involved in the EU-funded Ocean Data Interoperability Project, TERN (Terrestrial Ecosystem Research Network) is working with the US-funded DataONE (Data Observation Network for Earth) project, and the EMBL/EBI mirror at UQ is responsible for making the Platypus results available to Australian researchers more quickly by providing a local copy.

In all of these activities and the other eResearch infrastructure investments, Australia is aiming to be an international research collaboration partner of choice. We are putting in place a coherent environment for data storage, high performance computing, collaboration support through virtual laboratories and eResearch tools. ANDS is involved with all of this, as well as working with an increasing number of institutions who are committed to a data-intensive future. Our goal is to give Australian researchers an unfair competitive advantage on the world scene by making Australia the preferred collaboration location for data-intensive research.

"Researchers around the world sharing and using research data without barriers"

The Research Data Alliance

Andrew Treloar, ANDS

The main focus for ANDS' international activities over the next year will be the Research Data Alliance (RD Alliance). The vision for this new organisation is "Researchers around the world sharing and using research data without barriers". This international alliance came into existence through discussions within research funding bodies in the US and then moved to include the EU and Australia.

ANDS is the body in Australia who has been funded by DIISRTE to work on bringing the RD Alliance into existence. ANDS staff have joined a team from the US and Europe to form the RD Alliance—intended to be launched in March 2013—with early work on some of the data bridges in marine data and linguistics data already underway with our partners.

Further details at rd-alliance.org

The individuals currently working to bring the Research Data Alliance into being are:

- » Fran Berman, Professor of Computer Science, Rensselaer Polytechnic Institute
- » Juan Bicarregui, Acting Director e-Science, STFC Rutherford Appleton Laboratory
- » Leif Laaksonen, Collaboration Director, CSC Finland
- » Beth Plale, Director Data to Insight Center, Professor of Computer Science, Indiana University Bloomington
- » Andrew Treloar, Director of Technology, Australian National Data Service
- » Ross Wilkinson, Executive Director, Australian National Data Service
- » Peter Wittenburg, The Language Archive, Max Planck Institute for Psycholinguistics
- » John Wood, Secretary General of the Association of Commonwealth Universities

ANDS-funded projects

Edgar: mapping species and climate change

Marianne Brown, James Cook University

Climate change is affecting where flora and fauna are able to live. We are already seeing the environment suitable for the cassowary, being pushed further up the mountains of far north Queensland because of our warming climate. To conserve species we need to understand how their distributions are likely to continue to shift in the future.

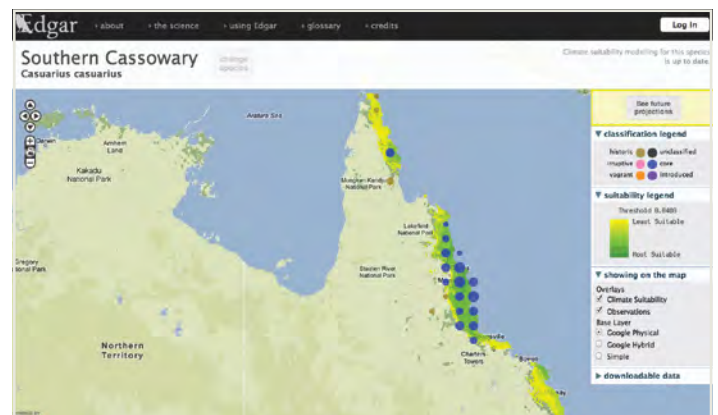
The Edgar website (tropicaldatahub.org/goto/Edgar)—recently launched by James Cook University (JCU) and Atlas of Living Australia (ALA)—allows researchers, the public, and policy-makers to see what climate change is likely to do to a particular species. This will assist them in making decisions regarding conservation and climate change action.

Edgar displays almost 18 million ALA bird observation records on a map. It then provides an interface for registered users to vet these displayed records by indicating whether they are likely to be correct or a mistake. The vetting information is then pushed back to ALA where it is included in their data store thus removing the need for re-vetting of the data for future projects.

Using the newly improved records, Edgar then automatically models reclassified data each time records are modified, and creates maps

of future projections of the species distribution. Edgar provides an animation of the species distribution changes from present to the year 2085 so that users can view the shift in species range under a range of climate change scenarios.

The team at JCU and ALA are using the current version of Edgar to discover what needs to be done in order to develop it to a global scale.



Screenshot of Edgar courtesy of James Cook University

The Gazetteer of Australia 2.0

John Weaver, Office of Spatial Policy and Greg Laughlin, ANDS

A gazetteer is a list of the place names, boundaries and physical features of the earth, and some information about each. It can be used for searching, illustration or disambiguation of names.

The Gazetteer of Australia 2.0 (gazetteer.mymaps.gov.au) is a new and improved version of the Gazetteer of Australia 2010, which was fundamentally a text-based product which could be downloaded from the website or interrogated through an online place name search.



Image of deserts in Australia courtesy of public domain NASA World Wind (CC BY 3.0)

The new Gazetteer infrastructure makes it much easier to get definitive and standardised location information both on the web and integrated programmatically into research applications, tools and data archives. It makes it easier to 'spatially enrich' research data, to connect with data from the same location and to address complex cross-disciplinary research questions based on location.

ANDS and the Office of Spatial Policy collaborated on the upgrade—which addressed issues of currency, discoverability and accessibility, and machine-to-machine capability sought by the research community. The Gazetteer of Australia 2.0 has the following features:

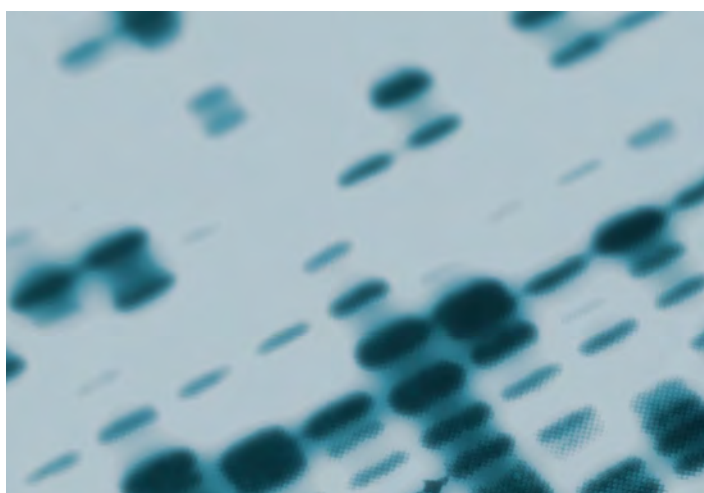
- » a web-based mapping service with the ability to graphically define search areas and display and download results
- » machine-to-machine functionality
- » an online, near real-time update capability for custodians
- » compliance with current international best practice guidelines and standards for gazetteers (WFS-G)
- » support for the integration of other Australian gazetteers including the Australian Antarctic Gazetteer and the Maritime Gazetteer into a single search environment

Event reports

Data Management webinars

Haidi Beard, James Cook University

The ANDS Data Management webinars are a well-organised series of presentations that relate to the exciting new field of research data management. Previous presentations can be viewed at leisure on the ANDS website (ands.org.au/presentations/audio-video.html), which is in essence, a centralised knowledge base for research data managers both nationally and internationally.



The webinars are presented by experts in the field and cover exciting new developments, as well as the standard components and best practices for good data management.

Although the content can be very specific it is easy enough to follow for a beginner. Usually I have a question to ask during the live webinars, and have found the responses to be very pertinent to my work. The webinars provide a perspective on current work in the field and a benchmark for your institution. By and large, effecting such a massive cultural change requires a group effort, and the collaborative and interactive nature of the webinars is an extra support that is most welcome. Thank you ANDS!

Getting around a table

Alan Glixman & Ingrid Mason, ANDS/Intersect

Collaboration will emerge when the environment is right, there is a means to enable it, a wider group of interested people and two people to start the process. The 'starters' in the New South Wales eResearch community for collaboration around metadata stores development are Vicki Picasso and Peter Sefton, from the University of Newcastle (UNEW) and the University of Western Sydney (UWS) respectively. There is a well-established alliance. Currently Vicki and Peter are involved in the development of the ReDBox application and they are actively collaborating in their

metadata stores projects. The wonderful spin-off of having these two in the mix in New South Wales along with a wider group of eResearch professionals all interested in information sharing means that 'many brains are making many light work'. ANDS has run regular metadata stores roundtables so people can get together and talk collegially. Stefania Riccardi from the Australian Catholic University says:

"To me the roundtables are 'moments of truth'. As the conversation spins and the speakers engage, I understand the real depth of the issues, priorities, relationships and not last attitudes. At the roundtable I get access to a wealth of "tacit" knowledge that emerges only when people sit together, talk without a script and release more information prompted by objections or the general mood of the audience. Listening and watching Amir Aryani talking about parties (NLA Party Infrastructure) made me understand the issues involved much better than any document I have read so far. Watching the genuine commitment of the people from UWS and UNEW, and the way they cooperate is stimulating. It is by listening to narratives and conversations among champions that I get motivation and focus."

No bow ties needed

Richard Ferrers, ANDS

To further build community, ANDS commenced an informal monthly meeting—Victorian (Tas.) Informal—that is designed to bring together ANDS partners and guests from IT, Library and Research Offices to discuss eResearch and Data Management, and to share knowledge and experiences. The Informal has a 'No minutes. No PowerPoint. Just stories' approach, as suggested and agreed by the inaugural Informal participants.

This approach frees up participants to share both good and bad experiences, and to avoid constraints typical when formally representing their institutions. Particularly well-attended sessions included chats with UK Digital Curation Centre visitors, and parting reflections from Monash eResearch Centre and VeRSI staff heading to sunnier climes in Brisbane. Victoria University is collaborating with ANDS on this ongoing-monthly event by hosting the event at their city location.

"...effecting such a massive cultural change requires a group effort..."

Queensland has its days!

Beth Crawler, University of Sunshine Coast

Queensland ANDS partners, gathered in Brisbane at Griffith University, Southbank on 8 and 9 August for the ANDS Qld Community Day and Data Citation Roundtable. Those involved in Seeding the Commons projects had the opportunity to hear reports from mature projects where ANDS funding has made a significant impact on research practices. The Community Day provided an insight into the diverse ways institutions have addressed research data management, through the Library, Research Office or IT, through self-service or mediated curation.

The Data Citation Roundtable showed the extent of cultural and structural change that data citation carries with it. It's a new

world for us all. Kim Finney reported on the established practices of the Australian Antarctic Data Centre while Michelle Teis from TERN challenged all of us to consider critically how well we are addressing data citation.

For those of just entering the world of data citation and management, the two days provided not only information about current good practices but the chance to consider ways to address the questions raised from the experiences of other institutions.

We are immensely grateful for the honest insights offered by so many of the presenters.

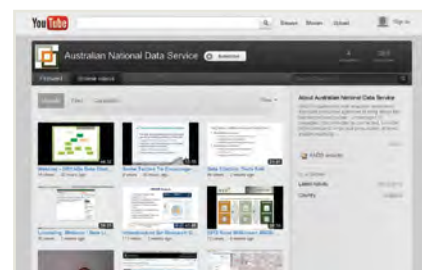
In brief

ANDS is now on YouTube!

New videos are added regularly and feature ANDS partners, international experts and ANDS staff. Current videos include the Data Citation series, the Australian Gazetteer, Licensing, Infrastructure for Research Data Management at Oxford University and more.

All ANDS recordings are available at: ands.org.au/presentations/audio-video.html

You can subscribe to the ANDS YouTube Channel to hear about new videos as they are uploaded. Find us by typing "ANDS youtube" into your search engine or youtube.com/user/andsdata



IP principles and data use

ANDS recently submitted feedback on the draft National Principles for Intellectual Property Management for Publicly Funded Research Conducted in the Public Sector, to the Commonwealth Committee on Innovation.

Our submission puts the case that there are significant benefits and opportunities in exploiting research data by licensing it, especially using CC BY, and placing it in repositories such that it can be reused by other researchers, government and industry, rather than protecting data for individual benefit.

The ANDS submission is available for download: ands.org.au/presentations/index.html

New syndication

Researchers now get more bang for their registration buck thanks to a recent agreement between ANDS and the Ex Libris Group, which will see registered collections descriptions in Research Data Australia syndicated through the Ex Libris portal, Primo Central.

For ANDS this syndication agreement is part of a strong agenda to disseminate the descriptions of research data that our partners provide us. We have similar syndication arrangements with other institutions, including the National Library of Australia, and actively expose the information to Internet search engines such as Google.

For more information visit: ands.org.au/news/ands-and-exlibris.html

EU opens access to data

The European Commission has adopted two important policy papers: on completing the European Research Area—a single market for research and innovation in Europe, and on opening access to scientific information funded by the EU. The aim is to boost Europe's innovation capacity and to address major societal challenges more effectively.

Read the full press releases on the two policies:

European Research Area: eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0392:FIN:EN:PDF

Open Access to Scientific Information: europa.eu/rapid/pressReleasesAction.do?reference=IP/12/790&format=HTML&aged=0&language=EN&guiLanguage=en

Forthcoming events

eResearch Australasia 2012: emPower eResearch

ANDS is hosting a booth at this year's eResearch Australasia conference. Come along, meet the ANDS team, chat to your Client Liaison Officer and join us in celebrating 'Better Data'!

ANDS will also be involved in a number of sessions:

Monday 29 October, 2012 – 1:40pm to 2:00pm	The Australian National Data Service – outcomes and achievements after 4 years David Groenewegen, Adrian Burton, Cynthia Love, Andrew Treloar and Ross Wilkinson
Monday 29 October, 2012 – 2:55pm to 3:55pm	User-facing Data Services and Capability Building – Institutional Development Teula Morgan, Lyle Winton and David Groenewegen
Monday 29 October, 2012 – 2:55pm to 3:55pm	Sustainable eResearch Software Development Rodney Harrison, Peter Sefton, Steve Androulakis, David Flanders and Steve Bennett
Tuesday 30 October, 2012 – 9:45am to 10:05am	The Proteome Browser – A Software Resource for the Chromosome-Centric Human Proteome Project Robert Goode, Anthony Beitz, Jeff Christiansen, Ed Nice and Ian Smith
Tuesday 30 October, 2012 – 2:35pm to 2:55pm	National Infrastructure Programs Panel Philip Tannenbaum, Glenn Moloney, Paul Sherlock, Nick Tate, David Toll and Ross Wilkinson
Tuesday 30 October, 2012 – 2:35pm to 2:55pm	Linking the EMBL Australia EBI Mirror to RDA Dominique Gorse, Pierre-Alain Chaumeil, Jeff Christiansen and Mark Ragan
Wednesday 31 October, 2012 – 11:10am to 11:30am	Identity Awareness: Toward an Invisible e-Infrastructure for Identifying Data and Authors Amir Aryani and Adrian Burton
Wednesday 31 October, 2012 – 1:30pm to 1:50pm	Data Citation: Stories from the Trenches Adrian Burton, Kim Finney, Anne Stevenson and Karen Visser

For more information on these and other sessions at eResearch Australasia visit: conference.eresearch.edu.au

Developers events at eResearch Australasia

In an eResearch first, ANDS, NeCTAR and RDSI are joining forces to host three events designed to get developers, designers and their managers involved in the conference as well as providing them with the opportunity to hear about other projects, network and share ideas.

The events include:

- » The Challenge
- » The Hardware Hack event: Internet of Instruments
- » The Developers Lounge

To register or for more information on these events please visit: conference.eresearch.edu.au/eres2012/developers-lounge

Get notified about our forthcoming newsletters via RSS feed: ands.org.au/newsletter

For more news, alerts, announcements and discussion subscribe to the ANDS General Google group by emailing: contact@ands.org.au

To learn about the upcoming ANDS events visit: ands.org.au/events

Join our conversation on Twitter: [@andsdata](https://twitter.com/andsdata)



ANDS Office
Building F, Level 6,
900 Dandenong Road,
Caulfield East, VIC 3145
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 (akdesign.com.au)

ANDS Project Partners:



MONASH
University



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