share
the newsletter of the Australian National Data Service
May 2012 : issue 12

MANAGING | CONNECTING | DISCOVERING | REUSING
RESEARCH DATA

# Collecting our Research Data

Cynthia Love, ANDS

Business has increasingly seen value in collecting market data. The Bureau of Meteorology is collecting water data to describe how the whole of Australia uses water. A botanist at the Australian National Herbarium can access all relevant research literature on banksias because it has been collected. Research data deserves to be collected too.

As the concept of data publication and sharing gains traction, and the volume of data available in the Australian Research Data Commons increases, the value of taking a collections approach emerges. This approach has long been used in libraries to introduce focus and applicability into the presentation of material. It can take many forms and track content across time series, origin, locations or related subject. Just as one would collect in the area of literature or artefacts, 'collecting with intent' becomes an important activity in data management. Determining what constitutes a collection is a vital component of this activity and ANDS has developed a guide: ands.org.au/guides/defining.a.collection.html

There are a number of collections approaches. Commonly it will arise from a single research project or from a single strand of research. Treating the data as a collection enables effective management, connection, discovery and reuse. A collection can support collaborative research activity where the data is brought together for a specific purpose on a discipline, national or international scale. It can also be in the form of 'reference' collections that are definitive, comprise specimens or observations and may be seminal in nature. Their function may be likened to reference material in libraries and be of broad uptake outside a specific discipline. A collections approach has the value of being able to showcase the data produced by an institution, either as research output or as support for Australia's research effort.

However, how we collect in Research Data Australia is also important in this regard. Research Data Australia (researchdata.ands.org.au) reflects Australia's data holdings and how we organise those holdings means that we can take a cross sectional view and create sub-collections. Thus we could bring together and profile the nation's data holdings on a given topic, or a variety of data collections that reference a specific location. The types of data could well cross discipline boundaries and provide a variety of perspectives. So, for example, Research Data Australia holds collections of marine data sourced from a number of institutions, which are then aggregated in the discipline specific Australian

## Inside Issue 12 – Collecting Research Data




ISSN 1836-9731

Oceanographic Data Network (AODN) collection. The value-add is the focus that this brings to connecting these collections to profile this research effort. These collections may also come together to provide pictures of research in other ways such as institution or specific research activity. Another example could be through the use of location. A slice can be taken through Research Data Australia that brings together all the data about a given location. Thus we could build a picture that shows the pattern of settlement, environmental observations, agricultural activity, the ecosystem and so on to give a more complete far-reaching view and enable new questions to be addressed.

Applying a collections approach to the management of data provides the following benefits:

- increases discoverability and browsability
- provides a more complete picture of data on any given topic
- shines a spotlight on important and significant data
- enables new and more complex research questions to be addressed

By working with collecting institutions, ANDS can leverage their understanding of the data and research environment in which it is created and used, to develop the descriptions that enable it to be presented in meaningful contexts.

# Executive Director's report
# Collecting with intent

Ross Wilkinson, ANDS' Executive Director

In my field of information retrieval research, the area was substantially changed when the U.S. National Institute of Standards and Technology introduced a text retrieval collection used by a large proportion of researchers in the field; it enabled new problems to be explored and provided a benchmark for testing new approaches to retrieval. It also led to a 100 per cent performance gain as researchers across the world saw what worked and what did not.

Increasingly organisations are establishing significant collections of research data that enables them to be placed at the centre of a research field. I like discussing with Deputy Vice-Chancellors of Research the "research data ambitions" of the institution. The variety of ambitions indicates just how research data supports the broader research agenda of the institutions. By way of example, James Cook University has a particular focus on tropical research, so it is not surprising that it wishes to provide a focus for tropical research data.

Disciplines also organise significant collections: at the recently launched Australian National Corpus of language hosted at Griffith University, researchers described the ability to not only access the consolidated data, but the value of tools that enabled new analysis. Astronomers are increasingly concentrating on gathering sky surveys, rather than highly focused observations, again enabling new sorts of questions to be answered.

Some of the most important data collections in Australia are being created by our infrastructure partners, such as the creation of a vastly enhanced ocean observing capability through IMOS, which is creating an internationally valuable collection of ocean data that is being used in a wide variety of investigations. As described elsewhere in this issue, the National Imaging Facility is creating a Mouse Brain Map that is again, internationally significant.

The significance of these collections is that it not only enables a research group to address new problems because of the scope of the collection, it also provides a reason for establishing international partnerships—sometimes it is a "chip on the table" that ensures Australia is vitally engaged in strategic research.

Sometimes a collection is a bridge that enables researchers across different disciplines to work on large problems together. Through its Applications program, ANDS is supporting initiatives that bring important data to the investigation of climate adaptation. This includes downscaled data being generated by Professor Andy Pitman's Climate Change Research Centre at UNSW, as well as local government data being gathered by Professor Brendan Mackey at Griffith University and field observations being gathered by Professor Steven Williams at James Cook University. This combined data environment will enable Australian researchers to address key problems for Australia. That's why it is good to collect data with intent.

*"Increasingly organisations are establishing significant collections of research data..."*

# Spoken like an Australian

Tim Thwaites, Science in Public

You bewdy! The more than 22 million of us on this island continent communicate in a language that is different from anywhere else in the world. Our English has been subject to a unique mix of environmental influences, other languages and dialects, and different social situations. And some of us use indigenous languages, sign language or community languages.

The Australian National Corpus (www.ausnc.org.au), launched on 26 March this year, has drawn together a wide range of samples of written and spoken Australian language for research purposes—and it will continue to expand. The samples come in the form of published texts, audio files and transcripts, and even video files showing the visual aspects of communication. In the future, they will also include electronic files of emails and blogs, and sign language.


Wordle created at wordle.net

"It makes existing collections—put together using public funds—widely accessible," says Project Sponsor, Dr Michael Haugh of Griffith University. "Bringing them together creates a much larger set of data on Australian language that people can search across and use."

The establishment of the Corpus was driven by linguists, applied linguists and language technologists to support collaborative research on what makes our language distinctively Australian. Already, it has contributed to a project on Irish influences on Australian English.

Australia is at the forefront of standardising such a collection of text, audio and video samples and making it available, Dr Haugh says. "There were ethical challenges with some of the older collections where samples of speech were recorded without permission. In some cases the communities were so small that the individuals speaking could be recognised."

# Documenting the human capacity for thought

Tim Thwaites, Science in Public

More than 2,000 of the world's 6,000 different languages are spoken in Australasia, the islands of the South Pacific and Southeast Asia. About 900 are spoken in New Guinea alone. But in this century the number of Asia-Pacific languages is likely to drop to a few hundred.

About a decade ago a group of linguists, ethnomusicologists and sociologists became concerned that important data was being lost. Significant recordings of language and culture from the 1960s and 70s, made on reel-to-reel tapes and cassettes, were becoming increasingly inaccessible due to deterioration of the media or lack of equipment to play them. In addition, the researchers who collected these recordings were retiring, sometimes dying, without being able to lodge their material in a place where it could be properly preserved for posterity.

The result was the establishment in 2003 of the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). "It helps us get a handle on the world's linguistic diversity, and also on human cognitive capacity in general," says Associate Professor Linda Barwick, a musicologist from the University of Sydney who is the Archive's Director.

The material contributed to the Archive is digitised and stored as standardised text, audio and video files together with metadata descriptions which can be searched. In fact, as part of the Open Language Archive Community, PARADISEC (paradisec.org.au) has participated in developing international standards for describing and managing ethnographic data. The Archive is interoperable with about 40 other similar collections worldwide. As part of the process of digitisation, a copy of the resulting file is always returned to the researcher, and where possible to the community in which the recording has been made.

Image: Vanuatu courtesy of Sarah M Stewart (CC BY)

# Where do Greater Gliders live?

Tim Thwaites, Science in Public

It began life in 2006 as a means by which the biological information stored in herbaria, museums, universities and other collections could be brought together and made available to a wider community. The prime movers were the heads of those information stores who put in a bid for funding to the Australian Government. Since then the Atlas of Living Australia (ala.org.au) has grown to be much, much more. In fact, one of the most significant contributors, Birds Australia, was not even involved originally.


Screenshot of Atlas of Living Australia

"The idea was to unlock the billion-dollar investment in developing biological data collections at a time when the information they contain is more important than ever," says CSIRO entomologist Dr John La Salle, now Director of the Atlas. "Over the next two to three decades we're going to face a number of major changes—growing more food on the same amount of arable land, maintaining sustainable ecosystems at a time of global change, dealing with emerging diseases—and our responses all require knowledge of biology and biodiversity."

The Atlas was founded on core principles of collaboration, open access, and dependable quality of data. The result is a collection of more than 30 million records not only of specimens from museums and herbaria, but also field sightings from amateur naturalists. Each observation, however, meets rigorous international standards for biological data. There is no charge for using the Atlas, and it provides tools for assembling and analysing the data in different ways.

The wide variety of people now accessing Atlas information—researchers, public interest groups, government agencies and commercial environmental analysts both in Australia and overseas—is proof of its breadth of appeal.

# Providing information where it counts

Tim Thwaites, Science in Public

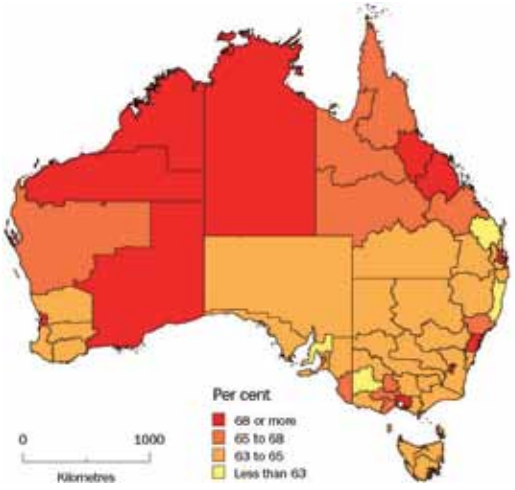
You don't need to ask the Australian Bureau of Statistics (ABS) why they collect data. After all, it is their job—mandated in two Commonwealth Acts of Parliament. "In this country, and in all democratic countries, statistics are regarded as very important sources of information for informed decision making," says Dr Siu-Ming Tam, First Assistant Statistician. "The legislation provides for the independence of the Statistician. There is also a legislative obligation for the Statistician to publish and release statistics once they are collected."

In 2005, after much debate, the ABS (abs.gov.au) was given Government support to make its information available free of charge, and it is overwhelmingly in electronic form. Although this has resulted in the loss of between $3 million and $4 million of revenue, Dr Tam says that for every dollar the ABS lost, the country earned about $4 through making use of the information, as highlighted in the Cost and Benefits of Data Provision report (ands.org.au/resource/houghton-cost-benefit-study.pdf)—commissioned by ANDS and conducted by Professor John Houghton of Victoria University.

Since free electronic access was introduced, Dr Tam says, "The extent of use of our information has grown exponentially." And it has also done wonders for the image of the ABS.

Before 2005, people used to stand up at conferences and criticise the Bureau for charging for information, particularly if they had contributed to it. "Now that our information is freely available, we often get standing ovations at conferences, and it gives us a lot of kudos in the community. And we receive much more cooperation and support for our work."

In 2008 the ABS introduced an open licensing system for its statistical releases by using the CC BY licence of Creative Commons.




Graph of working age population (15-54 years), Statistical Divisions, Australia 30 June 2010 courtesy of ABS

# Storing data makes it accessible

Tim Thwaites, Science in Public

Image: LaRDS tape storage library courtesy of Monash Medicine Multimedia

We have all heard the horror stories—a hard disk destroyed by fire, a USB stick left in an internet café, a disastrous computer crash before critical data could be stored. The consequences can be catastrophic—studies aborted by the loss of years of collected data, governments and corporations embarrassed by missing files, even national security put at risk. The remedy is sound data storage.

But there is another side to data storage, the legacy of high-performance computing. Data is now generated so quickly and in such volume that, without huge capacity to capture and store it, vital information is either lost or has to be discarded. The proposed Square Kilometre Array, the world's largest radio telescope, is expected to generate more data in a single day than the world does in a year at present. The volumes of data are measured in petabytes (billions of megabytes).

And in the case of nationally significant data collections or data generated from national facilities, such as Australia's radio telescopes, it makes sense to have a national capacity to store and maintain the information as well as provide the pathways of access to it. That was the thinking behind the $50-million Research Data Storage Infrastructure (RDSI) project, led by the University of Queensland and announced in the Federal Government's 2009-10 Budget.

RDSI (rdsi.uq.edu.au) is not being established in a vacuum, however. Some large institutions, such as Monash University, have already invested in enterprise-wide research data storage capacity, and have developed considerable expertise. The Monash system, known as the Large Research Data Store (LaRDS), has been in operation for about four and a half years, and the benefits are already apparent, says Professor Paul Bonnington, Director of the Monash e-Research Centre (monash.edu.au/eresearch). "Among other things, it improves the way people go about research and raises research performance. Because it facilitates reuse and sharing of data, the University's research has more impact nationally and globally, and Monash is now well recognised for the hosting and organisation of research data."

Where possible, Monash research data is described and stored in a central repository automatically as it is created. "Our aim is not to lose a single piece of data...ever. The system must be foolproof. So important research data is stored at different locations, and on two types of media—tape and spinning discs. Several times now, we have been able to restore files—that researchers thought were gone forever—because of a lost laptop or USB stick."

RDSI is still being established. It will comprise six primary data nodes, all joined by a dedicated high-speed network, together with up to four additional nodes which may be smaller and connected at slower speeds, says Project Director, Dr Nick Tate, who is based at the University of Queensland. Monash has put in a joint-bid with the University of Melbourne to host one of its data storage nodes in Victoria and is committed to helping make the RDSI project a success.

The first successful bidders will be announced soon, as will a first panel of preferred vendors of equipment. In the meantime a program to identify and fund suitable data services and collections is underway and the RDSI project is working with the Australian Academic and Research Network, AARNet (aarnet.edu.au), to develop the necessary high-speed network to carry the projected petabytes of data.

# A collection of important collections

Jeff Christiansen, Cynthia Love and Julia Martin, ANDS
Tim Thwaites, Science in Public

When the people of Harden, 125 km northwest of Canberra, decided they wanted to revegetate their area with indigenous species, they needed first to find out what those species were. They were able to do this by electronically interrogating the Australian National Herbarium data collection from home. This is just one example of how opening up Australia's data collections through the Australian National Data Service has led to an unexpected, positive reuse of information they hold. Some of the more significant Australian data collections are:

**CSIRO's Australian Biological Collections** (csiro.au/Organisation-Structure/National-Facilities/Australian-biological-collections.aspx) were established early in the organisation's history to assist research, and became national collections in the 1970s when they were gazetted by the Australian Parliament. They include among others:

- Australian National Wildlife Collection, Canberra
- Australian National Insect Collection, Canberra
- Australian National Fish Collection, Hobart
- Australian National Herbarium, Canberra

Most of the knowledge of Australia's flora and fauna comes from these collections. They support taxonomic, genetic, agricultural and ecological research into Australia's biodiversity within CSIRO (csiro.au) and other institutions, as well as industry. They are available through a number of portals such as Research Data Australia (researchdata.ands.org.au), CSIRO's Data Access Portal (data.csiro.au) and discipline portals such as the Atlas of Living Australia (ala.org.au).

Image: Kingfisher courtesy of the CSIRO Bird Biological Collection

The data contained in the **European Molecular Biology Laboratory Australian Mirror** (emblaustralia.org/Facilities/EBI_Mirror.aspx) of the EMBL European Bioinformatics Institute collection covers the 250 million DNA sequences published over the past 30 years found in the European Nucleotide Archive (www.ebi.ac.uk/ena). It also includes genomic sequences and accompanying annotations about gene activity for more than 60 organisms with backbones sourced through the Ensembl project (ensembl.org). In addition, the Mirror contains protein sequences and functional information found in UniProt (uniprot.org), a comprehensive, global, high quality and freely accessible resource. Making these large datasets available locally to Australian researchers allows faster access to the data as well as the option to copy the data and use it in intensive research.

Image: Clownfish courtesy of Shek Graham (CC BY)

Australia's **Integrated Marine Observing System** (IMOS) was established in 2007, and has successfully deployed a range of observing equipment in the oceans around the continent. The collected data is freely and openly available through the IMOS Ocean Portal (imos.org.au). It is used by the Australian marine and climate science community as well as international collaborators.

A good example is in the picture built up of the recent La Niña event through multiple observations—extreme events on the east coast, freshening of the Indonesian Throughflow current, and a marine heatwave down the west coast.

IMOS has been able to identify nearly 400 projects, publications and conference presentations using its data during the past 12 months.

The **Brain Mapping National Resource** project will bring together several important data collections concerned with mouse brain anatomy. These include high-resolution Magnetic Resonance Imaging (MRI) images of mouse brains and many other neuroanatomical data resources that have been produced through the Australian Mouse Brain Mapping Consortium (AMBMC). AMBMC (mousebrain.monash.org) is a national network funded through the National Health and Medical Research Council (nhmrc.gov.au) Enabling Grant Scheme with nodes at Monash University, the Howard Florey Institute, the University of Queensland, and the

Prince of Wales Medical Research Institute. The data collections will be pivotal in allowing quantitative comparison between multiple biological specimens—diseased and non-diseased, for example—which can be used in medical research.

The **Human Chromosome 7 Proteomics** project is building a system that will integrate information contained in several important international data collections concerned with proteomics, the study of protein structures and functions. The tool is being constructed for use in the international Chromosome-focused Human Proteome Project (C-HPP), where Australia has been assigned Chromosome 7. It will integrate information from at least eight major international proteomics and genomics resources including: neXtProt (Switzerland); Human Protein Atlas (Sweden); Peptide Atlas (US); the Global Proteome Machine (Canada); and Gene Cards (Israel). The project will be critical in the prioritisation of work for the C-HPP effort, providing a global view of existing proteomics resources.

**Geoscience Australia** (GA) is the nation's leading authority and a world leader in assembling geoscientific information and knowledge. This information is used by government and community to make decisions about the:

- exploitation of resources
- management of the environment
- safety of critical infrastructure
- wellbeing of all Australians

All the above activities are critically dependent on the data collections of GA (ga.gov.au). ANDS and GA recently collaborated on a project to create an enterprise-wide metadata profile for the GA data collections.

GA is continuing a data management program that will incorporate the rest of its data, as well as ensuring new data products are described and therefore able to be discovered by means of an electronic search.

# Chair's report
# Avoiding the tragedy of the commons

Ron Sandland, ANDS' Steering Committee Chair

This issue of *share* is focusing on the ANDS Collections approach. I am sure there is enough written elsewhere in the newsletter on the importance of this approach for me to simply say that it is a very significant piece, in the impressive national infrastructure that is ANDS.

But what is the totality of ANDS and why is it important?

In addition to the National Collections program, ANDS comprises some seven programs: Frameworks and Capabilities—which seeks to ensure that research institutions have the capability, systems and structures to effectively manage their research data; Seeding the Commons—to support researchers to capture, manage, share and reuse their data; Data Capture—to ensure that research data are physically able to be stored in an appropriate way; Metadata Stores—to complement the physical storage of data by providing the descriptions that actually make the data useful; Public Sector Data—to provide feeds of data from relevant public sector agencies which are often a critical part of the data researchers need to bring together to carry out their research; ARDC Core Infrastructure—to make the infrastructure operational; and Applications—to provide the tools and facilities to enable data to be analysed, interpreted and fully utilised.

ANDS has always focused unabashedly on institutions. For one thing they provide the local capability and infrastructure necessary to connect with the global infrastructure; for another they provide a home for and nurture the researchers who have the ultimate responsibility for the creative use of the data in their research.

The data revolution is well upon us. Our ability to gather data from a variety of instruments is increasing exponentially and our ability to exploit it to create new knowledge is desperately trying to keep pace with it. This revolution will affect institutions and researchers in profound but different ways. And this in a nutshell, is the reason the (now) Department of Innovation, Infrastructure, Science, Research and Tertiary Education (DIISRTE) invested in data infrastructure in the first place. But another aspect of their decision was to recognise this as an area of market failure and to avoid the "tragedy of the commons". According to Wikipedia, itself a part of the data revolution, the "tragedy of the commons" is a dilemma arising from the situation in which multiple individuals, acting independently and rationally consulting their own self-interest, will ultimately deplete a shared limited resource, even when it is clear that it is not in anyone's long-term interest for this to happen." Put simply, national investment is needed when there is no means of coordinating the investments of the research agencies into a coherent whole with a value that is significantly greater than the sum of the parts. ANDS seeks to provide mechanisms to avoid the "tragedy of the commons".

Globally the value of ANDS-type investments is increasingly being recognised as of immense value. A recent visit to Europe for a conference on Research Infrastructure clearly identified ANDS as a major global player—I have this from non-ANDS sources!—and research data will increasingly be the currency of international research collaboration. The ANDS Steering Committee is deeply committed to working with the leaders of ANDS, government and research institutions to ensure its value is understood.

# International engagement

Andrew Treloar and Karen Visser, ANDS

With international engagement playing an increasingly vital role in the function of ANDS, during the first few months of 2012 a number of ANDS representatives have been furthering these engagements. Ross Wilkinson and Andrew Treloar travelled to London and Copenhagen in March; Sarah Jones and Martin Donnelly from the Digital Curation Centre made the trip over from the UK to visit ANDS' Canberra and Melbourne offices; and Liz Lyon from UKOLN co-hosted a well-attended workshop in Melbourne with Andrew Treloar.

## European engagements

In London, Andrew helped run the International Summit on Digital Author Identifiers (DAI), which was run by the Knowledge Exchange (knowledge-exchange.info). The summit brought together a range of international experts in the field of author identifiers to discuss current and emerging issues. There was a realisation, across the board, that the different author identifiers—Open Researcher and Contributor ID (orcid.org), International Standard Name Identifier (isni.org) and the Virtual International Authority File (viaf.org)—are actually complementary, not competitors. This helped reinforce the willingness of the three organisations to talk about closer collaboration and possible alignment. Presentations and tweets on DAI summit are available on the Knowledge Exchange website (knowledge-exchange.info/Default.aspx?ID=62&M=News&PID=147&NewsID=140) and a concise report will be out soon.

Ross joined Andrew in London, to meet with Rachel Bruce of JISC (jisc.ac.uk) to discuss how to continue aligning research data infrastructure activities between Australia and the UK. JISC has some fantastic projects at the institution level, however it is missing some of the national infrastructure that ANDS has built here in Australia. It is clear that there is much to be gained from maintaining ongoing dialogue with our international counterparts, especially as we are all trying to solve the same data problems.

At the Data Access Interoperability Task Force (daitf.org) workshop in Copenhagen, discussions were held about global data infrastructure with many agreeing that the time is right for a global initiative on research data infrastructure. Australia is working closely with the North Americans and Europeans on how to improve international co-ordination, with ANDS continuing these discussions during a visit to Rome in April.

ANDS also made up part of the strong Australian contingent represented at the first International Conference on Research Infrastructure (ereg.me/ehome/31679/50320). There was a lot of interest in Australia's approach to research infrastructure and an even stronger emphasis throughout the conference on data. It was evident that the Australian Government's investment in eResearch infrastructure has provided a more holistic approach than many other countries, which gives Australian research a world leading advantage. Climate Change Research Centre Co-Director, Andy Pitman, has an interesting blog post (www.ccrc.unsw.edu.au/news/news/2012-03-27_reflectionsoneresearch.html) about the International Conference on Research Infrastructure from a researchers viewpoint, highlighting this advantage.

## Digital Curation Centre visits ANDS

In early April, Sarah Jones and Martin Donnelly from the UK's Digital Curation Centre (DCC) visited ANDS to start a dialogue about sharing ideas, resources and strategies to further data management in both Australia and UK. The DCC (dcc.ac.uk) is a world-leading centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management across the UK's higher education research community.

Their visit was full of planned sessions, starting with a one-hour virtual workshop for ANDS staff, followed by a one-hour webinar for data managers around Australia, which was so popular that a second was scheduled. These sessions looked at the DCC and the state-of-play of data management in the UK, which generated discussion about the DCC's non-contract institutional engagement model of 300 hours *in situ* support, the adaptability of the DCC Data Management Planning tools and their willingness to share the code.

In Melbourne, Sarah and Martin attended the second Victorian eResearch Informal Get-together where participants from Latrobe, Deakin, RMIT, Swinburne, Victoria and Melbourne Universities chewed over research data management, intellectual property, archiving, curation, and what to keep and what to delete. This was a great insight for ANDS partners into the world of data management in the UK.

Sarah and Martin expressed great interest in the strategies ANDS uses to overcome the tyranny of distance to communicate and engage with our community. Also of interest, was Research Data Australia going Open Source—it can be downloaded from: ands.org.au/resource/techdocs.html

Sarah provides some interesting insights into the visit to ANDS in her blog post: dcc.ac.uk/news/ands-dcc

## Community Capability Model workshop

One of the exciting challenges in data management, both locally and internationally, is how to best develop and support communities of practice. This is one of the drivers for the Developing a Community Capability Model for Data-Intensive Research project. This is a collaborative activity between UKOLN (ukoln.ac.uk) and Microsoft Research Connections (research.microsoft.com/en-us/collaboration), building upon the principles described in the book *The Fourth Paradigm* (research.microsoft.com/en-us/collaboration/fourthparadigm)

In February, ANDS assisted UKOLN's Liz Lyon to facilitate the fifth workshop in an international series, consulting on the draft whitepaper describing the proposed Community Capability Model (communitymodel.sharepoint.com/Pages/default.aspx), and receiving feedback from a range of practitioners. As well as helping facilitate the session, ANDS invited representatives from its partners to attend the session and offer their feedback on the Model. The discussion was lively with lots of suggestions for improvements to the Model. Interestingly there was a general consensus that it is hard to build sustainable communities of practice on short-term funding rounds, however there is an increasing number of university eResearch service providers providing a possible base for staff across a succession of projects. Importantly, an international 'buy-in' to solutions improves the size of the community and the solutions' sustainability.

# ANDS out and about
# Report on completed projects

## AusStage

Andrew Williams, ANDS

AusStage is a national collaborative project designed to provide an accessible research facility for investigation of live performance in Australia.

AusStage (www.ausstage.edu.au) endeavours to create a comprehensive listing of live events with dramatic content from 2001. It also extracts records for events dating back to the 1940s from archives and libraries of numerous theatre companies. The AusStage collection currently holds descriptions of over 65,000 live performance events and over 95,000 performers, as well as records for venues and performance organisations.

Image: Domain Theatre SA courtesy of Randy Larcombe

Flinders University is the current lead agent for AusStage and recently completed a Seeding the Commons project. The project made the AusStage dataset available in a flattened, machine-readable, standards compliant format—Research Description Framework—to support data reuse by researchers. The project also developed a process to regularly update the RDF dataset with the latest additions from AusStage collaboration partners Australia-wide.

AusStage was initially funded through an ARC (Australian Research Council) grant, and ANDS has subsequently funded significant developments to the resources AusStage provides to researchers. A number of enhancements were developed in a NeAT project co-funded by ANDS and the Australian Research Collaboration Service (ARCS), which include:

- mapping functionality—enabling visualisation of live performance through time, on a Google map
- network functionality—enabling exploration of networks of artistic collaboration
- audience research functionality—allowing capture of audience feedback in real time using mobile devices.

More information on the AusStage project is available here: beta.ausstage.edu.au

AusStage records in RDA: services.ands.org.au/home/orca/rda/search#!/q=ausstage/p=1/tab=All

# OzTrack

Leesa Clausen Brown, ANDS

The ANDS-funded OzTrack project, led by Prof Jane Hunter and managed by Dr Nigel Ward (University of Queensland), addresses the scale and complexity of animal tracking datasets and aims to accelerate scientific research by creating a common approach to the management and analysis of such datasets.

The OzTrack (oztrack.org) project has developed a set of tools that will provide the animal tracking community with simple access to data management, geospatial visualisations and statistical visualisations.

The ability to predict the movement of animals (such as crocodiles, sharks, whales, birds, bats, lizards, cane toads, koalas, dingos as well as production livestock such as cattle and sheep) has a key role to play in environmental and marine conservation and management. It is also of critical importance to addressing environmental challenges including invasive species, infectious diseases, climate and land-use change.

Descriptions of data collections using these tools are available:

The University of Queensland Dataspace:
dataspace.uq.edu.au/agents/2#

Research Data Australia:
services.ands.org.au/home/orca/rda/search#!/q=oztrack

Image: Crocodile with tracking device courtesy of OzTrack

# NeAT project—ASSDA Services for e-Social Science (ASeSS)

Leesa Clausen Brown, ANDS

The Australian Social Science Data Archive (ASSDA)—now known as the Australian Data Archive (ada.edu.au)—can help Australia's policy makers find the answers to questions as diverse as whether climate change can influence election outcomes, what are the details of immigrations rules, or whether child support payments should be introduced.

The NeAT project funded new software and tools that enable researchers to access forms of data not previously covered in the archive, bringing together a richer collection of social science data.

ASSDA's Executive Director, Prof Deborah Mitchell, says the NeAT project has changed the way ASSDA works, "NeAT has taken us from clunky technology and manual handling of data to smooth, automated and seamless web access". This new process provides an easier method for contributions to be made to the collection, with much improved support for the process of curation, regardless of the location of the curator.

An example of this was another NeAT project the National Criminal Justice Research Data Network (NCJRDN). This project addressed one of the biggest challenges for criminal justice researchers in Australia, which is that there is no single repository where criminal justice data is held. After gathering information on the relevant datasets across the country, NCJRDN approached ASSDA to host the data repository and portal.

The means that NCJRDN is providing data sets to ASSDA data curators, who then generate searchable metadata in a standard format and also ingest both the data and metadata into the repository. This results in better managed, connected, findable and reusable data sets.

Image: Kristen Granata, ASSDA data archivist courtesy of Randy Larcombe.

# ANDS events reports

## Licensing webinar

Margaret Henty, ANDS

The very first ANDS webinar was held on Thursday, 22 March. The topic was licensing, and our guest was Baden Appleyard, National Programme Director of the Australian Governments Open Access and Licensing Framework (AusGOAL).

The webinar took the form of a question and answer session. About sixteen people tuned in to the webinar and took the opportunity to ask questions about a range of topics. The most popular topic was Creative Commons: what happens when the CC version changes? Have there been any cases involving CC around the world, which might have an impact on the way it is used in Australia? When is the next version of CC due?

This Licensing Webinar was the first of four planned for 2012, with the next one to be held on 24 May.

Other webinars on different topics will be held during the year. Keep your eye on ANDS Events calendar: ands.org.au/events

## ANDS/NeCTAR developer days

David F Flanders, ANDS

*"A future where researchers and their teams can more easily reuse ANDS data and NeCTAR tools, as shared national research infrastructure will assure Australia will continue to lead the world in interdisciplinary research..."*

In support of this ANDS and the National eResearch Collaboration Tools and Resources (NeCTAR) project recently held the first of a series of workshops for developers and tech-savvy researchers to demonstrate how data and tools—in the cloud—can be easily reused.

The events not only presented how to use data and tools in the cloud, but also took the extra step of holding workshops in the afternoon, which enabled all participants to launch a tool and then put data into it. Best of all, because these tools are up in the cloud, they can be taken home to show researchers, thereby demonstrating the time savings in being able to trial new tools without having to buy a new server or transport a data hard drive across the country to reuse data.

Over 150 people attended the first two events in Melbourne and Brisbane and further events are being planned for Hobart, Perth, Adelaide, Canberra and Sydney. Keep your eye on the ANDS Events calendar: ands.org.au/events or the NeCTAR website: nectar.org.au

## ReDBox community day

Melanie Adamson, ANDS

Representatives from 14 universities, three state eResearch organisations and ANDS attended the ReDBox Community Day, held on 16 March. ReDBox (www.redboxresearchdata.com.au) is a system that provides the ability to describe research data, and make these descriptions (metadata) available to national/global registers. Some of the attendees were already using ReDBox, while the majority are considering its use.

Simon Pockley (ANDS) talked of focusing on developing communities, which enable beneficial connectedness, and how ANDS can use its birds-eye-view position to enable this. Projects can benefit from reuse of solution components, such as, database implementation.

Other key discussions that took place throughout the day were around NLA Integration, Metadata import (including RIF-CS), Research Management System Interfaces and embargo dates.

Although the ReDBox project wraps up in June, users of ReDBox will be able to receive technical support, for a fee, beyond the life of the project.

## Applications projects community days

Stefanie Kethers, ANDS

The first of the ANDS Applications Projects Community Days saw 27 participants from 15 institutions come together to present their projects, discuss a variety of issues, and learn about ANDS and how it can help them achieve research data outcomes.

The event, which was held in late March, revealed a number of unexpected connections between projects at the level of outputs, tools, and data sets. For example, there is potential for some of the Climate Change Adaptation projects to reuse image-viewing technology developed in the medical domain, or usage of the same spatial data sets across multiple projects.

Over the course of the event participants saw how they can draw on each other's expertise, and many conversations that started during the day have continued beyond the end of the event.

Feedback from participants has been very positive, and the Applications team is looking forward to working with them to help them bring data together to answer new and bigger questions.

# In brief

## Australian National Corpus launch

The ANDS-funded Australian National Corpus project was launched on 26 March. Australia's language data resources used to be scattered and relatively inaccessible at institutions around the country. The newly established Australian National Corpus (www.ausnc.org.au) has overcome this issue by establishing web application, built on top of a distributed technology stack, which combines existing corpora from partnering universities into a unified system. The system will be useful to linguists, applied linguists and language technologists, in encouraging new and collaborative linguistic research.

## TERN reuses ANDS software for data portal

ANDS has been working with the Australian Terrestrial Ecosystem Research Network (TERN) and developers from the Queensland Cyber Infrastructure Facility (QCIF) over the past few months to develop the new TERN Data Discovery Portal, which was recently revealed—alongside the new TERN facility portals—at TERN's annual symposium in Adelaide.

The TERN Data Discovery Portal will serve as the single entry point to TERN metadata for all levels of users—government, research organisations, universities, non-government organisations, professional bodies and the public.

The software (ORCA v7.1) used for the TERN data portal was developed by ANDS, and made available as part of the ANDS open source policy. ANDS is working with TERN to develop new search criteria fields to suit the type of searches required, such as searches that provide results for specific spatial areas.

For more on the TERN Data Discovery Portal visit: (tern.org.au/The-Australian-Terrestrial-Ecosystem-Research-Network-Data-Discovery-Portal-pg17727.html)

# Forthcoming events

## ANDS webinar series

ANDS is hosting a series of free webinars on Licensing and Data Management, which allows people working in research data management to engage with their community without having to leave their office.

The Licensing Webinar Series is in a Q&A style, where participants can engage in discussion about data licensing with Baden Appleyard of AusGOAL (ausgoal.gov.au). This series will run on Thursday, 24 May, 26 July and 27 September, 1-2pm AEST.

The Data Management Series features themed webinars addressing different aspects of data management. The series debuted with a webinar featuring the Digital Curation Centre UK. The second webinar, on 27 April, featured Julia Gross and Constance Wiebrands of Edith Cowan University, which discussed the outcomes of their ANDS-funded Project "*Seeding the Commons Data Management Plan and Policy*".

Places are limited, so register early to avoid disappointment. More information about ANDS events can be located here: ands.org.au/events

*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*
*Join our conversation on Twitter: @andsdata*

ANDS
AUSTRALIAN NATIONAL DATA SERVICE

**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 (akdesign.com.au)

**ANDS Project Partners:**

MONASH University
ANU THE AUSTRALIAN NATIONAL UNIVERSITY
CSIRO
Australian Government
Department of Industry Innovation, Science, Research and Tertiary Education