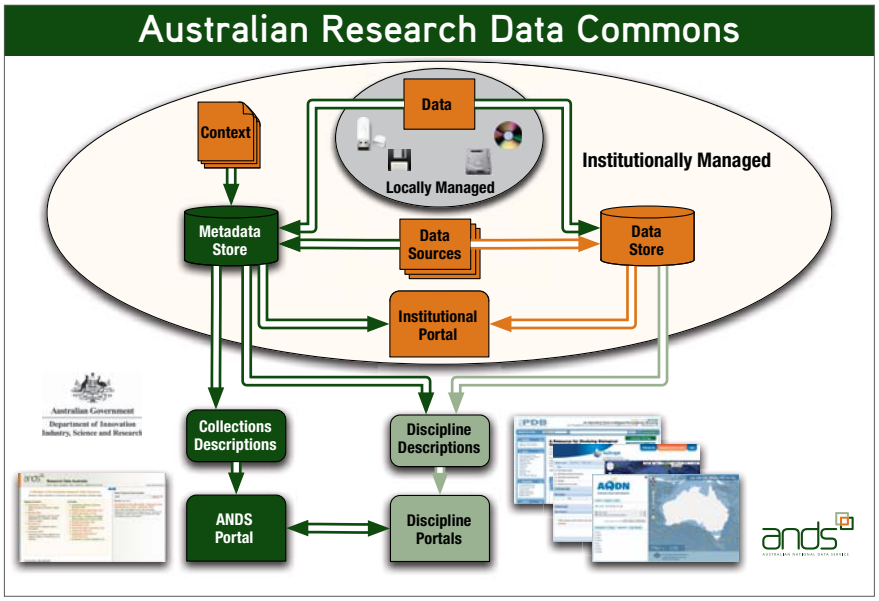
share
the newsletter of the Australian National Data Service
july 2011 : issue 09

Connecting our Research Data

Ross Wilkinson ANDS

Someone somewhere has data that could help your research. But you probably don't know it's there. Making data more accessible is central to our mission. We know that the more connected a piece of data is, the more likely it is that you will find the data you need when you need it.

Accessing and using research data can occur in many ways. Following a citation to a publication and then accessing the data might be one way. Finding a collection through a query using a task-specific portal, and determining whether the data was created using a compatible approach to your own data might be another. Knowing how a collection, a project and a research group are connected could be important, and each of these entities might support different ways of discovering a data collection.



The Australian Research Data Commons (ARDC) is a world-leading richly interconnected set of resources around research data that is being developed with the support of the Australian Government. The connections within the commons can be viewed and used within Research Data Australia (http://services.ands.org.au/home/orca/rda/), which is a window to Australian research data collections using connections to create enriched pathways. The commons might also be accessed through a national marine portal, through an international discipline portal, through the institution, or perhaps even by simply following links. The picture above indicates just how important connections are to the ARDC.

All of this only works if the ARDC is richly connected, and so the Australian National Data Service (ANDS) is working hard on enabling these connections. Within ANDS we are using an international standard (ISO2146:2010) model of data collections that recognises the importance of connections to context. We are gathering collections descriptions and associated descriptions.

The connections from collections to associated resources have to be authoritative, so ANDS is investing in services that enable these authoritative connections, such as those between people and places. ANDS recognises the high value of domain-specific access methods, and we are also working on building connections to facilitate access by others outside the domain. In this issue of *share* we highlight both international and national data initiatives with which we are working. Data associated with publications is increasingly seen as part of the whole of research output. By being able to publish and cite research data, the connections to publications are strengthened. Finally, we understand that connections between the communities who use the ARDC are crucial to its success. This issue of *share* is dedicated to the second of our four research data transformations: ***connections***, to achieve the goal of every path to research data being a productive one.

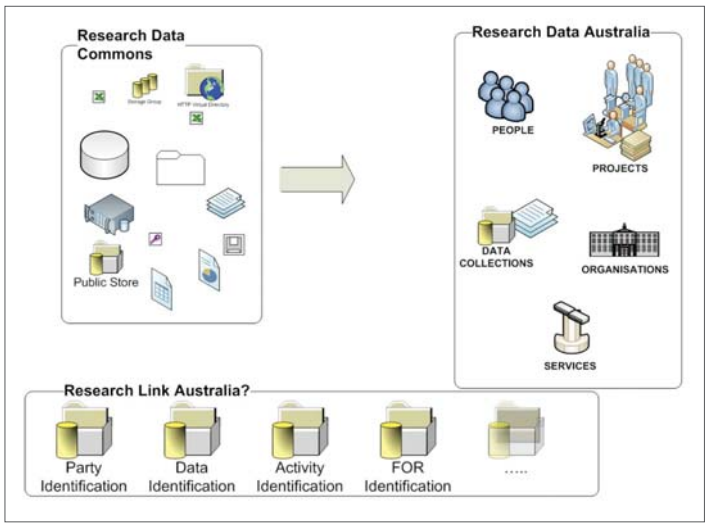
Inside Issue 09 – Special Issue on Connecting Research Data

ISSN 1836-9731

Enabling connections between people, places, projects and data

Pollyanna Sutton (Pollyanna is a freelance journalist), Andrew Treloar ANDS

Despite its name, the Australian Research Data Commons is not just about data – the context for this data is also crucial. The experience that the Australia National Data Service (ANDS) is seeking to enable within the Australian Research Data Commons (ARDC) through its Research Data Australia portal is less like a white pages service and more like interacting with Amazon.com. That is, the user can start with a page about the data, or about the researcher, or about the research project, and easily follow connections to more data from this researcher, or research project, or organisation. Instead of a single page, the user is offered a rich interconnected web of information that enhances discovery and provides valuable context.



Australian National Data Service (ANDS) data connections strategy (http://ands.org.au/guides/data-connections.html).

In order to enable this interconnected web, ANDS has been working with organisations and the Australian Government to develop infrastructure that will establish authoritative connections between people, places, projects and data. Dr Adrian Burton, from ANDS explained, "Because researchers are mobile, and design their own work, a professor may be involved in numerous projects around the country and indeed the world. The aim is to be able to find out not just what they have been doing, but to also link together the data they have created or have been associated with. ANDS hopes that by giving them a personal public identifier, and each research project a unique persistent identification code, everything that they are working on or that has been published becomes easy to find through association."

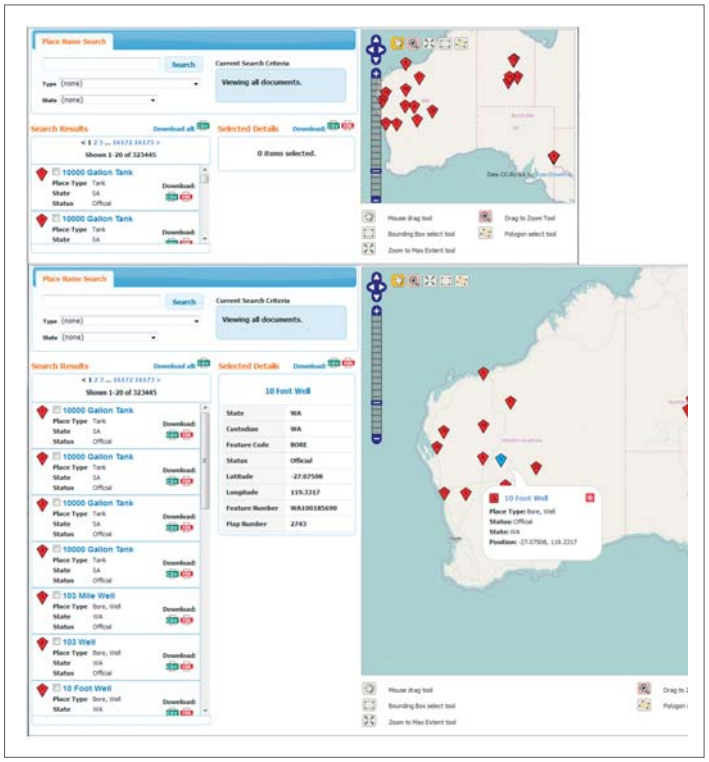
The National Library of Australia (NLA) has already developed an infrastructure for managing records about some categories of researchers and research organisations. The ANDS-funded Party Infrastructure Project is extending the use of this data throughout the Australian research sector and beyond. Users are already able to find some records about researchers and research organisations via Research Data Australia and the Trove People and Organisations zone (http://trove.nla.gov.au/people). These records link to the researcher's associated works thereby providing the public with improved access to Australia's research outputs. Additionally the Party Infrastructure assigns a public and persistent identifier to the researcher or research organisation. This identifier is associated with other known identifiers for the person or organisation and so the Party Infrastructure provides a public researcher identity service enabling the linking of resources across domains and services.

"This enables serendipitous connections across research activity."

According to Monica Omodei, from ANDS, in the future when cataloguing datasets, papers and other research outputs, it will be possible to reference researchers via a search lookup on a national researcher identity service rather than by typing a name. This ensures that all their research work over time in different institutions (and countries) can be linked.

Mr Basil Dewhurst from the NLA explained, "A key benefit of the Party Infrastructure is that it relates records about researchers and research organisations and their other associated identifiers. This means that a user or service could query the infrastructure with a known identifier and resolve it with another; over time these associations will encourage better management and curation of data generated by Australian research and supports the long-term access, use and re-use of this data." The project will conclude in July 2011 after which the research sector will be able to commence giving their researchers and research organisations public, persistent identifiers.

Another kind of connection is to place names. ANDS is working with GeoScience Australia (GA) to develop the Geoscience Gazetteer. This will provide an authoritative list of Australian place names and their geographical location. Through a free interface (http://www.mymaps.gov.au/gazetteer/), people can search for a place name and receive layered information about that place, such as how it was named, who named it, and what rivers and mountains are nearby. The information is returned on a map, which might look like a Google map, but rather than a static satellite picture, each location will have more detailed information that is sourced from Government databases.



Screen shot of the GeoScience gazetteer online service.

Dr. Greg Laughlin, from ANDS, explained that it is all about locating yourself in a space, for example a photographer might take a photo of a location and then use the Gazetteer to discover more information about the location. "The system has been designed to be elegant with some intelligence. Once users go through the Gazetteer, they can open up other datasets held by the Government which will begin to bring a place to life."

ANDS has provided funding for the Gazetteer and GA officers have spent time negotiating with the States and Territories to ensure that this information is now freely available. Dr Laughlin explained that GA has created connections to negotiate with the States and Territories to maintain and update their databases as new suburbs emerge. "It is a piece of wizardry, that will allow people to see information in a beautiful form live on screen with many layers," he said.

Another way of finding data collections is by subject. All Australian researchers are familiar with the Australian and New Zealand Standard Research Classification Field of Research codes. This is a standard classification scheme for research fields managed by the Australian Bureau of Statistics (ABS). ANDS is funding the creation of software infrastructure to publish these codes in a way that makes them available as connectors between related data collections. Users of Research Data Australia can easily select a subject code on any of the Research Data Australia pages to get a listing of other data collections and research projects in the same field. This enables serendipitous connections across research activity.

The last form of connection infrastructure is via research projects. The two major competitive research grant agencies in Australia are the Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC). Until now, the information about the projects they have funded has only been available through their websites as either text files or spreadsheets. ANDS is making this resource available in a way that enables connections to be made that provide greater context to a research project and its data. Thus, Research Data Australia allows a user to navigate from a data collection to the research project that funded it, and (in time) to the researchers associated with that project. It is this kind of flexible navigation and discovery that distinguishes Research Data Australia.

"Instead of a single page, the user is offered a rich interconnected web of information that enhances discovery and provides valuable context."

Connecting research to answer the big questions

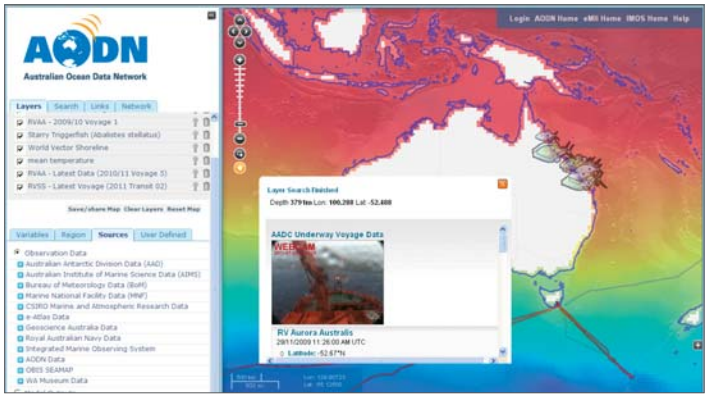
Pollyanna Sutton (Pollyanna is a freelance journalist), Cynthia Love ANDS

Gathering data about every aspect of the land and ocean is helping to build a more comprehensive picture of Australia and its surrounds, and also making stronger connections to enable 'big' research.

Providing the right infrastructure for this massive aggregation is just one part of the challenge. Another is connecting the data from differing sources to provide a comprehensive collection within a research discipline. While this collection enables discovery and reuse within the discipline it also enables connections beyond the discipline.

Dr Ross Wilkinson from ANDS explained, "The big picture is that we are trying to transform individual research data sets to data collections that are managed, connected, discoverable, and re-useable. Having data that is described properly is important because then other richer connections can be made such as which institution, project or person is responsible for the data. We have to connect discipline specific data access approaches to enable big questions like climate change to be addressed."

The Integrated Marine Observing System (IMOS) and AuScope have developed ways to bring this information together and make it available through their data portals, the Australian Ocean Data Network (AODN) Portal (http://portal.aodn.org.au/webportal/) and AuScope Delivery Portal (http://portal.auscope.org/portal/gmap.html) respectively. They can then feed this information to Research Data Australia and the Commons at large at different levels of complexity. Making it searchable and available on the web means that research teams across disciplines and around the world have access.



Screen shot of the AODN portal. Image courtesy of Jacqui Hope (University of Tasmania).

Dr Roger Proctor, Director of IMOS e-Marine Information Infrastructure (eMII) Facility, said "The AODN (Australian Ocean Data Network) is taking data from IMOS and commonwealth agencies with responsibilities for collecting marine data, and making it accessible through a public interface. This interface employs the same Open Geospatial Consortium (OGC) international standards as AuScope, enabling the wider community to access it. Even though we are working in a closed Australian community, the statistics show that people in China, India, Europe and South Africa are accessing the data".

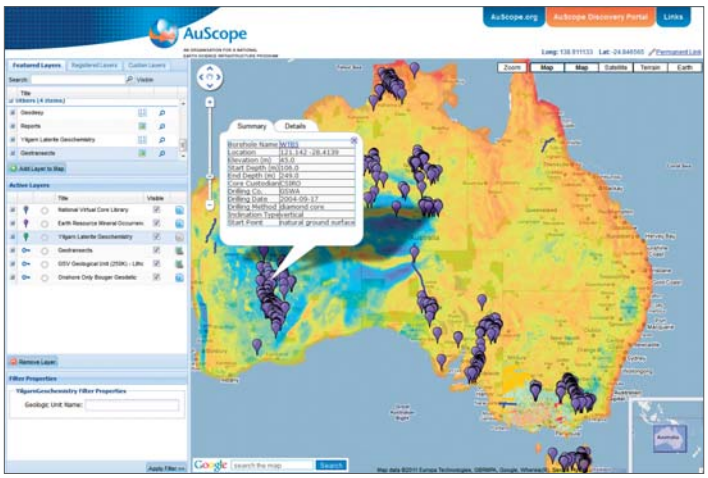
Dr Rob Woodcock, AuScope Grid Component Director, said one of the big challenges of providing an integrated service through the AuScope portal is connecting the data, because it is held across multiple government departments and institutions all with different software.

To resolve this AuScope is deploying a suite of tools called the Spatial Information Services Stack (SISS), which will help any research query in an area come back with a standard meaning. This enables data from different organisations to be discovered, connected, and integrated.

The SISS achieves this by supporting the minimum number of components of geoscience information that can then be transferred between different agencies. Because it operates on an international standard set by the OGC (the Open Geospatial Consortium - a group of 420 companies, government agencies and universities), it will make the data work better together.

Dr Woodcock explained the way it works is that a search is made and the question is converted to a private vocabulary, but then the answer comes back in a community vocabulary. For example, if a French person is looking for information on gold, they may type in the French word. The information may be listed in French, English, or by a number or symbol, the search query accesses all the associated information and returns it in a standardised form.

Dr Woodcock commented that everyone has the opportunity to contribute and build the information. There is support for third parties to contribute and individuals can also use the portal's webpage to discuss information.



Screen shot of the AuScope portal. Image courtesy of Dr Rob Woodcock.

Dr Wilkinson added, "If a researcher wants to explore the effect of below ground water flow on coastal water environments they need the deep access provided by the AuScope and the AODN portals, as well as the projects, the people and their connections that are visible through Research Data Australia – it's a great example of the value of connected, findable and integrated data enabling new research."

Mirroring the richness of Australian species

Pollyanna Sutton (Pollyanna is a freelance journalist), Andrew Treloar ANDS, Jeff Christiansen ANDS

The recent launch of the first ever European Molecular Biology Laboratory (EMBL) Australia mirror of EMBL-European Bioinformatics Institute (EBI) database in Australia will enhance access to genomic and other biomolecular data about Australian species and enable the Australian research community to undertake new cutting-edge research.

For the past 25 years, DNA and protein researchers around the world have been required to upload their datasets to one of three international facilities located in the UK (EBI), US or Japan when they publish work. These facilities then exchange and synchronise their data. This has created a massive resource and has also resulted in an exponential growth of the EBI archive: 2-3 datasets are uploaded every second and the archive currently contains 10 petabytes and climbing.

In order to continue to provide quality access, data safety, and capacity to the global research community the EBI has encouraged the setting up of mirrors around the globe. The first such Mirror has been developed in Australia, with the Australian National Data Service (ANDS) providing a portion of the funding. The EMBL Australia mirror of EMBL-EBI is located at the University of Queensland but is accessible nationally and internationally via (www.ebi.edu.au).

Professor Mark Ragan, Group Leader for Comparative and Computational Genomics at the Institute for Molecular Bioscience of UQ said, "The new system will allow researchers to access data, use it in a workflow space, or download it to their institutional super computers for use. It will also enable researchers to access data 10 times faster, reducing download times."

Information relevant to Australian researchers will be added to the EMBL Australia mirror of EMBL-EBI in multiple ways. Existing information captured by international researchers that relates to animals and plants from Australia will be tagged in the database. This includes over 100 collections of DNA and protein data that have been generated by Australian researchers, submitted to the main EBI site, reviewed by a manual curator, and then returned to the EMBL Australia mirror of EMBL-EBI. Computerised workflows run on the Mirror will also analyse existing data and generate new collections.

Software funded by ANDS takes the data in the EMBL Australia mirror of EMBL-EBI and provides descriptions of these significant collections to Research Data Australia. This process is exposing over 13,500

"2-3 datasets are uploaded every second and the archive currently contains 10 petabytes and climbing."

collections of DNA and protein data for different species and groups of animals and plants that have been identified as Australian in the Atlas of Living Australia. These descriptions go down to the level of individual species. This enables connections to be made with other bio-archives such as the Terrestrial Ecosystems Research Network. For instance, a researcher looking at a location may identify a species of interest, search Research Data Australia and find links to the data in the Mirror, as well as to related research projects, researchers and organisations. This enhances further discovery and gives the researcher an opportunity to determine how valuable this information is to their query.

The EMBL Australia mirror of EMBL-EBI will provide high-throughput local access to international molecular bioscience data and add value to international data that can be applied to Australian problems. It will also showcase selected Australian tools, provide advanced technical training, offer novel career paths, and engage user communities.

Through the international connections established, this project has created a wonderful opportunity to expand the Australian Research Data Commons by making this wealth of DNA and protein data discoverable to researchers in other fields around the world and visible through Research Data Australia.

Professor Ragan commented, "We will start with some of the data and some of the services, and over the next year or two we will expand, and work with EBI to develop new services, which will have global reach."

Connecting datasets to publications

Adrian Burton ANDS

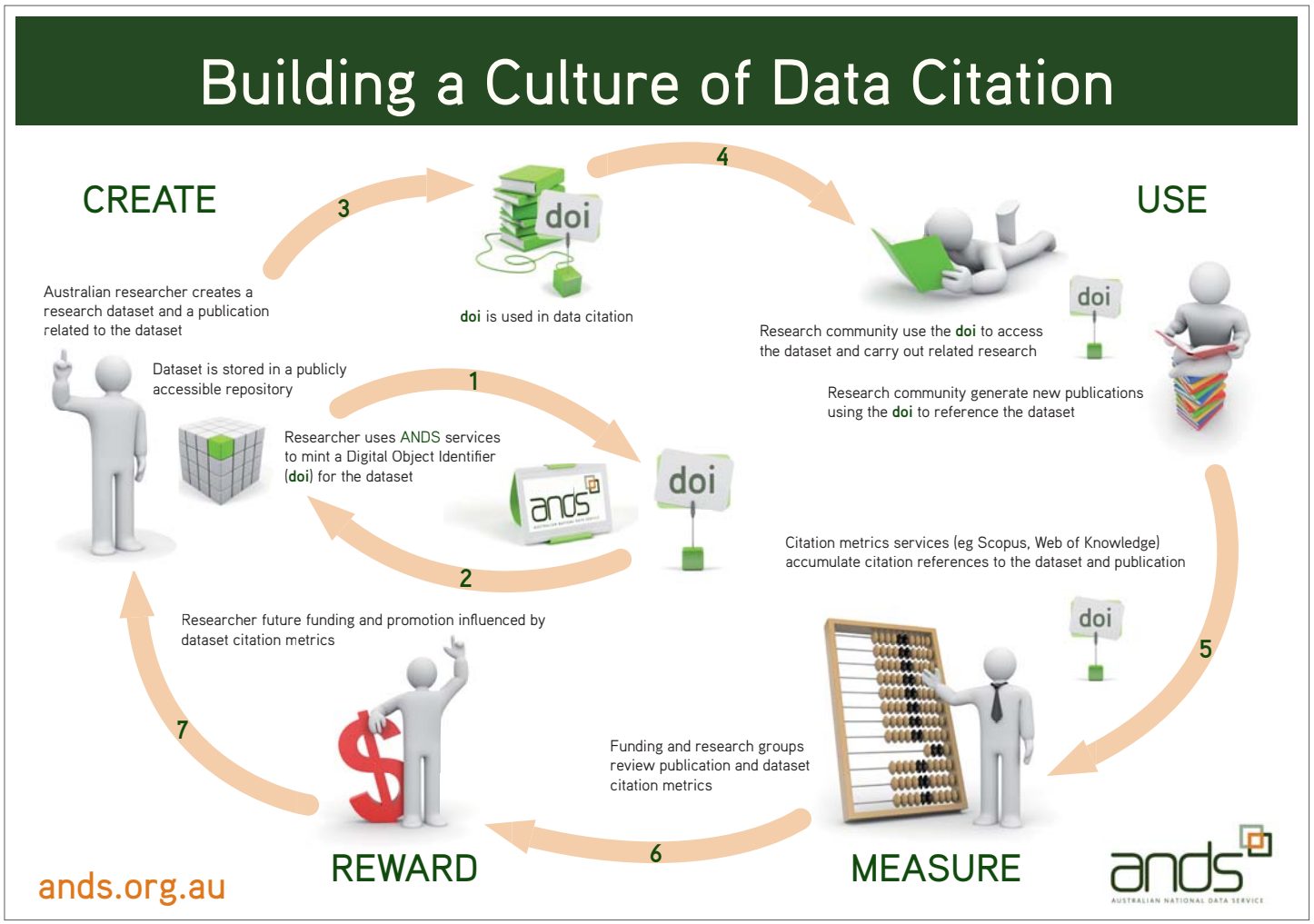
From July 2011, ANDS is piloting a Digital Object Identifier (DOI) service for research datasets. A DOI is a unique number that helps identify and keep track of objects on the internet, like a registration plate for a car, or ISBN number for a book. The Australian National Data Service (ANDS) service will allow you to allocate one of these numbers to a dataset.

Why? Unique identifiers are used to keep track of things – like bar codes on DHL packages allow you to automatically track the movements of packages. Journal publishers use DOIs to automatically track references and citations to journal articles. If you visit an online journal, more and more articles have a digital object identifier, and the official citation contains the DOI number. Systems track and keep count of references and citations to the journal article using the DOI number.

A similar system is emerging for published datasets. If they have a DOI, then references and citations to them can be tracked in the same way. This means acknowledgement for researchers and research organisations that publish data.

And the big publishing houses are coming on board. Recently, Elsevier started reciprocal linking – the automated linking of data sets deposited at the PANGAEA data centre to corresponding articles in Elsevier journals on its electronic platform ScienceDirect and vice versa. The DOI unique number allows that linking to be automated.

The DOI has another function: put http://dx.doi.org/ in front of any DOI number, and with one click you are taken to the current location on the internet of the object identified. What's more, the publisher of the dataset can update that URL location at anytime with the central DOI infrastructure, and clicking on a DOI link will now take you to the new address. That's potentially "farewell" to broken links and 404 errors on the internet, which is really important for the scholarly record where references would ideally persist for future generations of scholars.



The journal publishers originally adopted the DOI system partly for that very reason; so that if one publisher went out of business or was purchased by another, the DOI references to their articles would still work, because the URL locations of their journal articles could change and simply be updated in DOI system.

DOI is a technology that enables data citation, citation indexing, and persistent linking. But technology is only a part of the equation. Community citation standards, a culture of data citation, and professional reward are part of the culture change needed to make any technology reach its potential. ANDS is part of an international consortium, DataCite, which provides not only access to the global DOI infrastructure, but also an international advocate to support the culture change to data citation.

The DOI system enables persistent connections not only between datasets and publications but also between researchers and rewards. It may turn out to be the missing link of a virtuous feedback loop that makes the scholarly world go round.

Connecting data, connecting people

Margaret Henty ANDS

The two ANDS Boot Camps held in mid-2010 each began with a speed networking event. The participants lined up in two long rows and were given two minutes in which to introduce themselves and to meet the person opposite. The result was cheerful mayhem, with lots of chatter, a bell ringing every two minutes, instructions to move on to the next person and no-one feeling that they had been given long enough to really get to know the person they were talking to. At the end of it, one of the participants was overheard to say "Well, it was worth coming, just for that! Now I know who to talk to when I want to talk something over".

This is one example of the need for those who are undertaking ANDS projects to connect with each other and one of the many ways in which ANDS has responded to that need. The two Boot Camps each brought together twenty people from institutions around the country to learn about data management and sharing and to meet up with others new to the field. Those people soon came to realise also how important it was to be developing connections between different parts of their institutions. More recently, there have been Community Days held in Brisbane, Sydney, Melbourne, Adelaide and Perth, each one providing more opportunities for ANDS Partners to get together, discuss their issues and hear about the work of others.

ANDS provides other ways to make connections. The ANDS Partners Google group now has over 200 members from universities and other research institutions around Australia, all of whom are associated with an ANDS-funded contract of one kind or another within their institution and all of whom are engaged in establishing the infrastructure to support improved data management within their institutions. The ANDS Community Bulletin Board has a similar number of people registered to share documents and ideas, and non-partners can have access to much of the content. The ANDS website includes a registry of research data infrastructure and tools, containing information about particular partners, projects and people to help link up those with similar interests.

Creating an Australian Research Data Commons will involve the development of human networks as much as it involves all the complex technical infrastructure which lies beneath.

August 2010 boot campers (L - R) Caroline Norris (University of Wollongong), Vicki Picasso (University of Newcastle) and Caroline Drury (CAIRSS) participating in a marketing exercise.

August 2010 boot campers (L - R) Amanda Nixon (Flinders University), Toby Burrows (University of Western Australia), and Aaron Ballagh (James Cook University)continue building connections over dinner.

Community Event reports

VIC & TAS ANDS Community Event

Stefanie Kethers ANDS, Steve Bennett ANDS

The ANDS Victoria and Tasmania Community Event was held on the 28th of March at the State Library Victoria. Over sixty participants attended the day-long event, which brought together people working on ANDS-funded projects at more than 15 organisations across Victoria and Tasmania. In the morning, representatives from the various institutions presented on their Data Capture, Seeding the Commons, and Metadata Stores projects already underway, while the afternoon was devoted to breakout groups discussing project roadblocks, learnings and issues brought up by the ANDS Victorian / Tasmanian community. Topics discussed included engagement with researchers, automatic creation of RIF-CS/harvesting, and data management policies.

WA ANDS Community Event

Matt Wyatt ANDS

The ANDS Community Event in Western Australia was a showcase of ANDS-funded and NeAT projects, which demonstrated WA's excellence and diversity in eResearch. Held on the 8th April in Technology Park, the event was made up of speakers from Edith Cowan University, Murdoch University, Curtin University, University of Western Australia, National Criminal Justice Research Network, Australian Social Science Data Archive, Western Australian Ocean Data Network, AuScope and the Terrestrial Ecosystem Research Network.

Having such a high 'caliber' of representation available in a single room for an entire day proved valuable for discussion and collaboration going into the future.

Group activity tackling metadata management (NSW & ACT).

NSW & ACT ANDS Community Event

Neil Dickson ANDS

On the 17th & 18th May 2011 ANDS organised its biggest outreach Community Event to date in May at the Australian Technology Park, Sydney for ANDS clients in New South Wales and the ACT. The event was held over two days and was attended by over 50 people representing more than 16 universities or research organisations.

The program for day one focussed on information and understanding issues associated with Seeding the Commons projects and allowed mature projects to pass on important successful strategies and advice to newly started projects. There was ample opportunity to discuss a wide range of important issues during lively group discussion sessions. It was very noticeable how many people took advantage of the opportunity to introduce themselves to others over the breaks and at lunch.

Day two targeted Data Capture projects. This was also well-attended and everyone was equally enthusiastic. Once again a highlight was the willingness and determination by ANDS' clients to make connections and recognise opportunities to partner and collaborate.

Tackling an issue associated with implementing metadata stores (NSW & ACT).

ANDS Community Events: reflections on the first completed round

Angela Lang ANDS, David Groenewegen ANDS

ANDS has been running a series of Community Events in the first half of 2011, as part of our ongoing community building efforts. These have been held in Brisbane, Perth, Adelaide, Melbourne and Sydney, with a total of 193 partner staff in attendance. These partners presented on 92 different ANDS-funded projects.

Feedback from partners indicated that they found hearing about other projects, making connections with others, and learning from other partners and ANDS staff most valuable. Having various experts in the room also meant that partners were able to seek clarification on issues that were specifically relevant to their project. During each event, partners broke into discussion groups to either problem-solve or share issues. These discussions have been captured by ANDS to help prioritise future support.

To further build a community of data management experts, ANDS plans to host more events over the next two years, with a focus on connecting those with specific roles and needs in common, for instance for data librarians, for technical staff, for those interested in ontologies etc., and is interested in receiving further feedback. At the end of each event, partners were encouraged to join the ands-partners google group and the Community Bulletin Board, to help further share information. Details about these and other resources for partners can be found at (http://www.ands.org.au/partners.html). Notes from these discussions can be found on the ANDS community bulletin board at (http://community.ands.org.au/) (login required).

Chair's report – Ron Sandland

The ANDS Steering Committee has recently been focusing on the future of data infrastructure in Australia in response to the process the Department of Innovation, Infrastructure, Science and Research (DIISR) is undertaking to develop the next Roadmap for Research Infrastructure.

In our view ANDS has played an extremely valuable role in creating infrastructure that has enabled Australia's major universities and research institutions to understand the need for effective data management mechanisms and practices. The approach taken by ANDS appears to be somewhat unique and there is much anecdotal evidence of international recognition of its value. We believe that there is a national competitive advantage to be gained through the effective and efficient enablement of data management, including storage, access, sharing and use of data.

Future investment in research capabilities, such as those that emerged under NCRIS, will continue to play a very important role in the future. The importance of data has been stressed by many of the Expert Working Groups working with DIISR to define the priorities for future infrastructure investment. These include environmental sustainability; health frontier technologies – the "data deluge" safeguarding Australia, and; cultures & communities.

The approach taken by ANDS has been largely institutionally focused. We believe that this has been appropriate given the relative immaturity of understanding of data management and data sharing issues in the research community. The level of understanding of these issues or the availability of sustainable infrastructure or services has not reached the point where the most appropriate investment in research data infrastructure would be to smear it across the disciplinary capabilities. As well as investing in the infrastructure needs of these capabilities, we believe there is a need for a continuing investment in an overarching data management capability which is responsive to the needs of the individual disciplinary capabilities. Such a nationally co-ordinated infrastructure will enable:

- The most effective use of existing and future investments in research data infrastructure, positioning Australia as a focus of data-intensive research
- Research to be undertaken that would otherwise be impossible as the data resources would not be available
- Australia to maintain its reputation and capability as a globally leading player in data intensive research.

The real dividends from investment in research data infrastructure will emerge in the efficiency and effectiveness of the Australian research community, its ability to tackle new (system-level) classes of problems, its ability to rapidly synthesise data from a number of sources (including public-sector sources) and to tackle problems of national and global significance. It will be a critical engine for multidisciplinary research whose importance is becoming ever more evident. A sustainable investment in data infrastructure will also position Australian researchers to be members of elite research communities in which data sharing is, in effect, the entry ticket.

With appropriate investment in national infrastructure, the data-intensive science of the future will be an exciting place for researchers as the articles that appear in *share* already attest.

ANDS' international connections

Andrew Treloar ANDS

Joint Information Systems Committee (JISC) Managing Research Data (MRD) international workshop

This workshop was held in Birmingham on March 28 and 29, 2011. It was an opportunity to analyse and evaluate the outputs and progress of the JISCMRD Programme, as well as a strong international dimension which allowed sharing of approaches to, and models for, policy, strategy and implementation. This dimension reflected the key partnerships which the JISC in general, the JISCMRD Programme in particular and the DCC has been building through the IDCC Conference, the Knowledge Exchange and other initiatives. International participants were drawn from Australia, the US, Germany and the Netherlands. Andrew Treloar from ANDS delivered one of the closing plenaries on day one, looking at Data: Its origins in the past, what the problems are in the present, and how national responses can help fix the future.

From an Australian perspective, the two most notable things were the extremely high-quality of the MRD-funded projects, and the interest in the UK in the national services that ANDS is establishing.

(http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmevents/mrdinternationalworkshop.aspx)

Prato Invitational Research Data Infrastructure Workshop

In April 2011, the Australian National Data Service (ANDS) facilitated a three day invitational workshop at the Monash University Prato Centre (http://www.ita.monash.edu/index.html). The reason for choosing this venue was to facilitate the greatest possible participation from northern hemisphere invitees. The intention was to bring together research data infrastructure providers and stakeholder representatives in order to explore the potential for much closer collaboration and co-ordinated activity.

Day one involved a series of introductory presentations from the national representatives to ensure that all participants had a common understanding. Day two was focused on sustainability, as this had been one of the major issues arising from the pre-workshop planning discussions. Work was focused around how to ensure the sustainability of software development, service provision and resource creation. Day three then moved on to produce 27 concrete actions arising from the workshop.

These covered sharing of findings, roadmaps, and good stories, co-ordinated approaches to funders and disciplines, better aligned approaches to data management planning, work on the integration of all research outputs, making progress towards identifiers for everything, and further activity around collections registries (including the possible uptake of the Australian solution, and exchange between registries). ANDS' staff are already having follow-up meetings to keep the momentum going after the workshop, and to start to deliver on the outcomes.

Group photo at the end of the workshop: L to R Lynn Copeland (Canadian Association of Research Libraries), Paul Walk (UKOLN), David Shotton (Oxford University), Linda O'Brien (Griffith University), Simon Hodson (JISC), Hans Pfeiffenberger (Alfred Wegener Institute), Adrian Burton (ANDS), Stefan Winkler-Nees (Deutsche Forschungs Gemeinschaft), Herbert Gruttemeier (Institut de l'Information Scientifique et Technique), Marc Dupuis (SURF), Ross Wilkinson (ANDS), Andrew Treloar (ANDS), Alfred Heller (Technical Information Center of Denmark), Rachel Bruce (JISC), Peter Doorn (Data Archiving and Networked Services), Matthew Wollard (UK Data Archive), Paul Bonnington (Monash University), Kevin Ashley (Digital Curation Centre), Birgit Gemeinholzer (University of Giessen), Maurice Vanderfeesten (SURF).

Australia-EU Research Infrastructure Workshop

As an initiative of the Australian Government and the European Union, an Australia-EU Research Infrastructure Workshop was held in Brussels on 4-5 April 2011. The workshop consisted of three concurrent thematic sessions (involving researchers in synchrotron science, microscopy and ocean science) with cross-cutting plenary sessions providing context and integration. ANDS led a plenary session on Scientific Data Infrastructure in partnership with representatives from the European Commission, and other european organisations including the European Commission High Level Expert Group on Scientific Data , and the OpenAIRE+ project. This session preceded the individual breakout sessions and emphasised the importance of data issues, something that was picked up in much of the later discussion.

Following on from the workshop, Australia (through the agency of ANDS) and the EU (through the Directorate Generals of the Information Society and for Research and Innovation) have agreed to cooperate on research data, interoperability, access and global governance. Specifically, the cooperation will work to harmonise the requirements for research data plans as part of the process of applying for competitive grants in both jurisdictions. The cooperation will also aim to bridge key infrastructure initiatives addressing different scientific communities and organisations, promoting data interoperability and access. The intention is to have coherent approaches to, and requirements on, researchers in both jurisdictions to ensure greater international consistency.

Meet the ANDS Staff

Ingrid Mason

Ingrid Mason's role with ANDS is to provide outreach and support for the ANDS-funded projects in New South Wales, she is based at the Intersect office in Sydney. A key part of that work Ingrid finds satisfying is supporting the needs of researchers and helping to contribute to the wider goal to improve data management in accordance with ANDS aims – so there is value generated at multiple levels. Ingrid has a background in digital development in the galleries, libraries, archives and museums (GLAM) and university sectors in New Zealand and Australia. One of the benefits of having worked in these sectors is that she has an understanding of how the research materials and outputs are managed and a richer knowledge of the links to research practices and scholarly communication. In her work undertaken for ANDS a highlight has been to gain a stronger understanding of "data generating and seeking behaviour" of researchers undertaking data intensive research, using instruments and new technologies. A challenge associated with this insight into data intensive research has been the recognition that increasing the proportion of digitised materials as a data resource is only one means of enabling researchers in the humanities, arts and social sciences to participate as actively in this approach to research. In the coming months she aims to build greater awareness of the ANDS data connections strategy and how those data services can be utilised in the systems and infrastructure development in the ANDS projects being undertaken in New South Wales.

Julia Martin

Julia Martin is a Business Analyst with the ANDS Public Sector Data (PSD) team. She has previously worked at CSIRO and Australian Federal government agencies as well as the financial sector in the United Kingdom. Her ANDS role involves being a primary point of contact for a number of ANDS and National eResearch Architecture Taskforce (NeAT) projects; providing support to PSD partners in meeting their contractual obligations; plus reviewing, assessing and reporting on ANDS and NeAT projects. These projects include the AuScope, Spatially Integrated Social Science (SISS), CSIRO Water Resources and Australian Ocean Data Network (AODN) Underway projects. She also works on site at government agencies to develop the infrastructure for automated feeds of collection records of their data. A recent engagement was the successful GeoScience Australia activity.

The recent months have seen a number of PSD clients achieve major milestones as they provide collection records to Research Data Australia for the first time. Julia said that it is fantastic to witness the culmination of everyone's efforts. The next few months will see the PSD team commence new engagements with a number of Federal Government Agencies. Julia and the team expect to face some exciting and challenging times.

In brief

Calling all data scientists

What is a data scientist? According to CIO Magazine (http://www.cio.com/article/684344/The_6_Hottest_New_Jobs_in_IT?page=2&taxonomyId=3123), there is a saying that "data is the new oil", ready to be analysed and exploited. Enter the data scientist who can manipulate large quantities of data, take advantage of the opportunity to improve analysis and delve deeply to uncover patterns and oddities.

In Brisbane, a Data Scientist Meetup Group gets together each month in a convivial setting to explore issues of data management, data visualisation, open data and data science. Anyone interested can connect to (http://www.meetup.com/The-Brisbane-Data-Scientists-Meetup-Group/events/calendar/), to discover more.

Data-DOI on German E.coli dataset

During May and June this year a deadly bacterial outbreak in Europe of the O104:H4 E.coli strain led to over 42 deaths and more than 3,800 people falling ill. The genomic data was released under a Creative Commons licence to allow scientific teams from around the world to collaborate and contribute in an unprecedented manner to the public health efforts to control the deadly strain. The data set has also been given a Digital Object Identifier (DOI) so that it can be cited correctly. The University Medical Centre Hamburg-Eppendorf and BGI-Shenzhen who worked together to sequence the bacterium and assess its human health risk, took these steps to maximise the data's usefulness to the scientific community.

(http://dx.doi.org/10.5524/100001)

Rewarding a (treasure) Trove

For most people, the word Trove is associated with the word treasure, and the National Library of Australia's Trove service is proving to be just that.

Trove was recently awarded the 2011 Excellence in eGovernment Award, as well as the Service Delivery Category Award, at the Australian Government's ICT Awards presentation in Sydney on June 1. The Special Minister of State, Gary Gray, said of Trove that it "shows how Government is harnessing digital technology to improve service delivery, linking Australians to millions of resources available online, and in Australian libraries, cultural institution and research collections".

ANDS is proud to be associated with the further improvement of Trove. The Party Identifier Project being undertaken by the National Library with ANDS funding will provide considerable benefit to Trove and Research Data Australia by allowing the systematic identification of Australian researchers as they move from project to project and institution to institution.

ANDS extends its congratulations to the National Library for its exceptional work in developing this national treasure.

Sharing project – Griffith Metadata Hub

In the last issue of *share*, the Griffith Metadata Exchange Hub was featured in an article. In the wake of the ANDS-funded project's success, Griffith has published various reports and other magazine articles including a paper published by International Association of Scientific and Technological University Libraries (IATUL) as a result of the 32nd Annual IATUL Conference Proceedings. The paper titled 'Shared Benefits from Exposing Research Data' explores in greater detail the project, processes and the infrastructure development.

(http://hdl.handle.net/10072/39023)

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

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy program and the Super Science Initiative.

This newsletter is designed by AK Design (http://www.akdesign.com.au)

ANDS Office
C/- Monash University
Clayton, VIC 3800
Telephone: 03 9902 0585
Email: contact@ands.org.au

ANDS Project Partners: