

## Reusing our Research Data

Ross Wilkinson, ANDS



When Captain Cook sailed into Botany Bay on May 7, 1770 he measured the temperature of the sea. He then recorded this information in the ship log and was then stored by the British Admiralty. Sea temperature measurements are important information in understanding our climate. So such long range measurements enables richer models to be explored.

Australia invests heavily in capturing new research data—through very expensive instruments, such as the synchrotron or a telescope, or through very expensive capture of data by teams over long periods of time—sea sensors, flux towers, and surveys conducted by teams of researchers observing their environment or interviewing people. This investment is necessary to answer incredibly important questions—how should Australia respond to bushfire? how can Australia best use its water resources? what is happening to Australia's climate?

However some of this data is only available for the experiment that generated it. Sometimes this is because it is appropriate to do so to preserve privacy, but sometimes it is because not enough effort has been allocated to making the data available for further uses. Data is more valuable if it can be used by researchers:

- » at a later time to verify results
- » to examine new questions arising out of current investigations
- » to answer new questions

However, data is perhaps most useful when it can be combined with other data to enable new classes of questions to be explored.

A powerful example of this comes from the vision of the Human Variome Project that emerged from discussions lead by Professor Cotton and his team at the Genomic Disorders Research Centre at the University of Melbourne. The vision is no less than the improvement of human health through the worldwide collection and sharing of all genetic variations effecting human disease—this requires both clinical data and genetic research data. It is incredibly ambitious but arises out of the incredible advances that have occurred in genetic research that has taken place as a result of sharing genetic data. This was neither natural nor cheap. The Human Genome Project speaks to the power of sharing and reusing data.

What are the steps to achieving this, in *share*, we have over the last few issues described: managing data—deciding early what will happen to the data and taking steps to carry out these decisions; connecting data—to the researchers and the research: and discovering data—making the data findable by others. We now describe ways of supporting reuse, which includes appropriate recognition of others data through appropriate citation, appropriate rules of use though simple data licencing approaches, examples of reusing data, and the power of aggregated data to support new investigations.

By making Australian data reusable, we make the data way more valuable, and enable Australian researchers to collaborate over this data to answer new questions crucial to our future.

### Inside Issue 11 – Special Issue on Reusing Research Data

- |                               |                      |   |                       |
|-------------------------------|----------------------|---|-----------------------|
| » Reusing our Research Data   | » NeAT report        | » ANDS out and about—report on completed ANDS-funded projects | » Meet the ANDS staff |
| » Chair's report—Ron Sandland | » Conference reports |   | » In brief            |

## Many routes to opening up data

Tim Thwaites, Science in Public

James Cook University in Townsville is one of the very few developed nation universities in the tropics. Yet the tropics is where two-thirds of the world's population lives, and faces distinctive issues in health, agriculture, the environment and managing resources. That's why a team at James Cook University's eResearch Centre has been heavily involved in putting together a Tropical Data Hub ([eresearch.jcu.edu.au/projects/tropical-futures-a-tropical-data-hub](http://eresearch.jcu.edu.au/projects/tropical-futures-a-tropical-data-hub)), to draw together information on anything to do with the tropics across Australia, from the Kimberleys to Cape York.



Image courtesy of AURIN

The urgent and significant problems of urban areas, on the other hand, have more to do with managing people than resources—planning, transport, the impact of pollution and so on. So the data being put together and administered by AURIN (Australian Urban Research Infrastructure Network) at The University of Melbourne is much more likely to be about human beings, and that brings with it significant issues of privacy and commerce.

The two projects have more or less the same goals—to create a one-stop-shop of datasets, and a suite of tools with which to analyse and manage them. Ultimately, they both hope to expose quality Australian research to the world in ways to enhance uptake and reuse by managers, decision and policy makers, researchers, and the general public. They are roughly at the same stage of development—both aiming to be online in 2012. But their approaches are remarkably different because of the different nature of the data and the disciplines involved.

The Tropical Data Hub, for instance, initially is focusing on convincing people across Australia's north to deposit datasets with the eResearch Centre where they can be properly managed and made available to others. Much of AURIN's information, however, is more sensitive and comes from government departments and agencies, from universities and research institutes, and even from private companies, where it is already well looked after. "So we'll be using a federated infrastructure, we want to keep the data as close to the original custodians as possible," says AURIN's senior project manager, Dr Martin Tomko.

Both will develop and make available tools to analyse, visualise, store and share data. Early-adopters in the Tropical Data Hub have been working with the developers to build tools with an eye to incorporating information from the widest range of sources—including amateur naturalists and "citizen scientists".

In order to manage the project in an orderly way the AURIN team decided, after consultation with the research community, to develop tools to analyse the accumulated information through 10 "lenses" relating to particular groups of urban issues. In the first year, for instance, the project is funding the development of tools for three of those lenses: population and demographic futures; economic activity and urban labour markets; urban health, wellbeing and quality of life. The next two lenses will relate to housing and transport.

But the most significant difference in the approach of the two projects is one of access. Where the Tropical Data Hub is concerned with drawing together and making available information as openly as possible, this is not possible in the same way with AURIN ([aurin.org.au](http://aurin.org.au)), because of the sensitivities and legal constraints of dealing with human data, particularly health information.

Initially, users of AURIN will need to be registered with the Australian Access Federation ([aaf.edu.au](http://aaf.edu.au)), a system by which bona fide researchers, through the institutions to which they belong, can have their identity and credentials verified electronically to allow them access to sensitive information. Most Australian academics are automatically registered through their university email addresses.

In future, however, it is likely that tools will be developed so that non-Australians and non-academics can register, as well as to allow particular data sets to be open only under specific restrictions, or upon payment of a licence fee. "Some of the agencies are under legal obligation to protect their data in certain ways," says Dr Tomko. "We need to have a legal framework in place, otherwise we won't get access."



The Daintree Rainforest - image courtesy of Kyle Taylor

# Resetting the gauge for more efficient data exchange

Tim Thwaites, Science in Public

The efficient exchange of data makes things happen. Be it sharing mineral resources information or traffic accident statistics, we live in a data driven world. But often data gets stuck within government for reasons resembling Australia's rail network in the 20th century.

In 1901, Australia federated—in part to free up trade within the continent. But it wasn't until 1962 that the first train ran directly between Sydney and Melbourne. For the previous 79 years, passengers and freight had to change trains at Albury, on the NSW-Victoria border due to different railway gauges. The issue was not so much technical, but cultural—frustrated by politics and legal agreements.

Governments across Australia are responding to the increasingly vocal calls by the community for access to and reuse of government information. Five Australian governments have enacted laws to push more departmental information out into the community. But legislatively compelling departments to release more information doesn't always enable it to be freely used.

National Program Director of AusGOAL ([ausgoal.gov.au](http://ausgoal.gov.au)), Mr Baden Appleyard has been working for more than five years on the legal and policy issues of making government data more accessible and reusable. "We are now paying for our information management sins of the past." Baden says. "For example, agencies frequently sever government data from the agreements under which it was created. And it's not unusual for departments to be uncertain about how they can reuse that data, or whether they can even supply it to the community to reuse." Baden is talking about the various legal issues that apply to the collection of data, such as copyright, privacy, commercial-in-confidence considerations, the involvement of multiple people or organisations, and prior contractual agreements.

The Australian Governments Open Access and Licensing Framework (AusGOAL), provides solutions to these challenges, and is the one-stop licensing shop enabling governments to clear blockers to the reuse of public sector and publicly funded data. The Framework is equally applicable to data produced in universities and research institutes, and thus has been adopted by ANDS.

At the heart of AusGOAL are Creative Commons licences. "The AusGOAL policy deliberately favours the least restrictive licence appropriate to the material" says Baden, "because it gives the broadest ability to reuse the information or data." In most cases this will be the Creative Commons Attribution licence, known as CC:BY, which simply requires attribution of the data source. But there are seven other licences to suit specific circumstances. The next most popular, the Creative Commons Attribution—Non-Derivative, allows use, but not blending or changing data. Two of the eight licences are not based

on Creative Commons—a specialist software licence and a restrictive licence template.

Once data is licensed under the AusGOAL framework, everyone who wants to reuse it in future knows where they legally stand. As such, the development of AusGOAL has made a significant contribution to the work of ANDS, to open up information accumulated by public institutions or using public funds for the widest possible reuse.



Screenshot of [ausgoal.gov.au](http://ausgoal.gov.au)

And it makes a difference. The Australian Bureau of Statistics, an AusGOAL advocate that began freeing up access to its data in 2005, has documented growing community support for its efforts and an exponential increase in the reuse of its information since that time, says First Assistant Statistician, Dr Siu-Ming Tam.

The AusGOAL framework extends beyond government data, it is equally applicable to research data. There is a growing trend within the research sector to apply the creative commons licences to research data as it encourages data reuse through the provision of clarity around the permissions, term and conditions for reuse of the data. The main benefit of applying the same framework to research and government data is that it facilitates the combined use and reuse for more effective research. ANDS has invested in AusGOAL to ensure that there is no changing trains, when moving from Government data to research data.

# Show me the data: citing information increases science impact

Tim Thwaites, Science in Public

Fewer than 200 scientists are involved in the operation of the Hubble Space Telescope. This "inner circle" is responsible for deciding where the telescope points in the sky, and what information it gathers. Yet collectively these researchers have produced only about half the research papers based on Hubble data. The other half has been written by astronomers, reusing data stored and made accessible through the Hubble Legacy Archive ([hla.stsci.edu](http://hla.stsci.edu)).



Data Citation Leaf courtesy of Robert Cook et al, Environmental Data Science & Systems, Oak Ridge National Laboratory

It's a growing trend—one group of scientists undertakes the experiments or observations and makes the data available for their colleagues to interpret. It then becomes important for the source of the data in scientific papers to be acknowledged and accessible so that the real impact of organisations, or infrastructure such as the Hubble Space Telescope, can be seen.

That's one of several reasons for the growing trend of citing data sources in scholarly publications, says Dr Adrian Burton, Director of Services at ANDS. "ANDS has an agenda to enable better quality science by promoting the greater discovery of data, and acknowledgement of those who create and manage it. Citing of data promotes all those things."

Another benefit of citing data has emerged from the work of Dr Heather Piwowar, now a postdoctoral fellow at Duke University. She has found that papers with links to underlying data have an increased likelihood of citation and therefore a higher citation ranking and impact factor.

The datasets do not necessarily have to be published in journal articles or scholarly communications. They can be archived anywhere. Likewise, they do not have to be sets of measurements or numbers. Digital Object Identifier (DOI) codes can be allocated to spreadsheets, visualisations, algorithms, graphics, computer models, pictures and diagrams, sound files, videos or any product of scholarly or research effort.

Access to the data, says Dr Burton, also leads to better quality science because it allows the evidence to be checked, verified and, if necessary, bolstered or repeated. And the descriptive metadata associated with archiving the datasets for citation, not only encourages discovery, sharing and reuse, but also can be used to assert primacy of work, if necessary.

ANDS is working on two fronts to encourage data citation—through promoting a culture of data citation at research organisations and publishers, and by providing the latest electronic infrastructure to facilitate the process.

After a successful six-month pilot with CSIRO, NICTA and Griffith and Monash Universities, ANDS launched its Cite My Data service ([ands.org.au/services/cite-my-data.html](http://ands.org.au/services/cite-my-data.html)) at the eResearch Australia 2011 Conference. Research organisations, data centres and scientific infrastructure providers can register with Cite My Data, which provides them with the technology to allocate a DOI code to each dataset generated. This allows each citation of the dataset not only to be tracked, but also to be linked to the original information. Putting <http://dx.doi.org> in front of any DOI code, creates a URL which can be clicked to get to the current location on the internet of the object identified.

The entire system is at no cost or effort to individual researchers. In fact, for the most part, just like the assignment of DOI codes to journal articles or ISBN numbers to books, the process will remain unseen to the data creators.

ANDS is a member of the international consortium, DataCite, which not only provides access to the global DOI infrastructure, but is also an international advocate to support the change to data citation. At the local level, it is working with scholarly societies, publishers and research organisations to encourage the expectation of data citation and develop appropriate formats in which to do so.

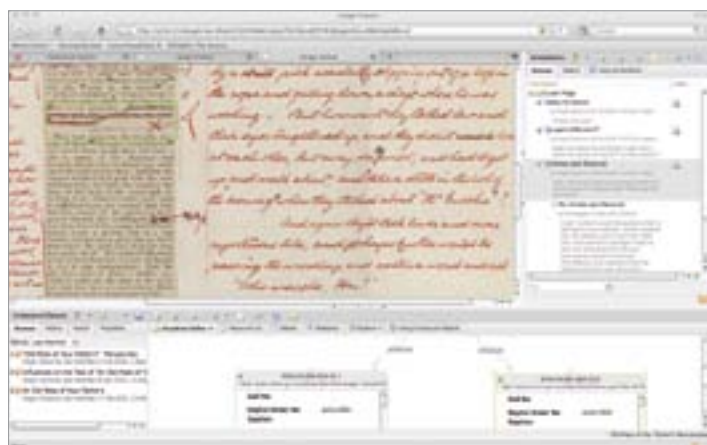
[http://dx.doi.org/10.1594/  
PANGAEA.484677](http://dx.doi.org/10.1594/PANGAEA.484677)

Example of a hyperlinked DOI code

# From controversy to classic: revealing literary undercurrents

Tim Thwaites, Science in Public

It's hard to imagine there were calls to censor Ruth Park's much admired novel *The Harp in the South*—about life in the slums of Sydney in the 1940s—when it was first serialised in the *Sydney Morning Herald*. Although the work had already won first prize in the paper's novel competition, it became evident through a storm of letters to the editor, that many readers felt uncomfortable with a depiction of their city that included drunkenness, prostitution and violence. Today the book is respectable enough to be included on secondary school reading lists.



Screenshot of the LORE tool

The community debate about the novel is an intriguing topic for academic study—revealing much about the social mores of the time. But this kind of research is traditionally time consuming and hard to share.

University of Queensland (UQ) literary scholar, Dr Roger Osborne is using an electronic tool known as LORE to construct the sequence of events and arguments associated with the serialisation. LORE, or Literature Object Reuse and Exchange, was developed by the UQ eResearch Lab for the AustLit community with funding from ANDS and the Australian Research Collaboration Service.

In future, LORE will enable Dr Osborne to trace textual variations both in the serialisation, and subsequent editions of the book, which has never been out of print since first published in 1948. The result is likely to be an electronic scholarly edition of the text to allow further literary investigation.

The study is an excellent example of the difference that the application of electronic tools can make in encouraging the reuse of information in the humanities. "In the past, researchers used to photocopy such resources in libraries and gather them together into a folder that went in a filing cabinet. This is an electronic way

of being able to deliver the same thing, in a form you can annotate and share with collaborators." Dr Osborne says.

LORE grew out of a discussion between the leader of the eResearch Lab, Prof Jane Hunter, and Ms Kerry Kilner of AustLit about the electronic tools that could be developed to analyse information about Australian authors and their texts held in literary databases, and then to visualise the results.

That interaction resulted in the three-year Aus-e-Lit project which has developed two primary tools—and made others available—to encourage reuse of the material on the AustLit website ([austlit.edu.au](http://austlit.edu.au)). Dr Osborne was able to assemble the material for his work on *The Harp in the South* by conducting a "federated search", trawling several other significant databases as well as AustLit—in particular the National Library of Australia's Trove database ([trove.nla.gov.au](http://trove.nla.gov.au)), which includes the original newspapers. From this search he generated a compound object or resource map of original texts, as well as commentary, criticism and bibliographic material.

The major difficulty encountered in the project, says Prof Hunter, was not the development, adaptation or acquisition of tools by the team, which was led by software engineer Ms Anna Gerber. "What is most challenging is encouraging the literary community to use what we have provided. Although the development of the tools was driven by user needs, some AustLit sub-communities find the technology intimidating. Roger Osborne's role as a community liaison person has been critical to encouraging the literary community to adopt the tools and provide us with feedback."



Screenshot of AUSTLIT taken in the Mortlock Chamber, State Library South Australia  
Photo Credit: Randy Larcombe

Prof Hunter and Ms Gerber are now intending to build on the LORE work to develop an eResearch tool that will enable researchers to compile and publish electronic scholarly editions of important works for future research.

# A peek into the future of DataCite: a chat with Jan Brase

Karen Visser, ANDS

Jan Brase, Managing Agent of DataCite attended eResearch Australasia 2011—eXtreme eResearch ([conference.eresearch.edu.au](http://conference.eresearch.edu.au)) to present as the closing speaker and at the ANDS DataCite *Building a Culture of Data Citation* workshop, which followed the conference. In a spare moment at conference, Jan agreed to chat about how DataCite ([datacite.org](http://datacite.org)) is working to further the use and acceptance of Digital Object Identifiers (DOIs) for research datasets.

One theme Jan explored in his presentations was the almost universal paradigm shift in understanding how tremendously expensive it is to produce datasets and how, up until recently, there has been little awareness of how much of research investment was lost when the resultant data was poorly managed.

Jan says: "...even the ability to cite data might not be enough incentive for the scientists to publish their data". In view of this, Jan is working towards offering DataCite services, which would provide much more comprehensive metrics than just citation counts. These services may include how often the DOI was resolved and/or downloaded or mentioned in social media such as Twitter and blogs, to give a much more holistic view of how

datasets contribute to the scholarly record. Another future service DataCite hopes to implement is a mechanism to access raw data to build up visualisation tools for data citation.

DataCite is not yet three years old but has already made remarkable progress in promoting the concept of data citation. Jan's vision for DataCite in the next three years is equally impressive:

- » Definitely millions more DOIs for datasets!
- » New services, such as the ability to cite persistent queries on datasets
- » Capitalising on the opportunities of metadata stores
- » Increased coverage, filling current gaps such as Eastern Europe, Russia, South America and Africa

DataCite is critically interested in the role ANDS plays in ensuring the value of data citation is recognised within the Australian research data community. The ANDS Cite My Data Service using DOIs forms a major strategic initiative in supporting Australian research data discovery and reuse.

Jan's DataCite blog can be found at: [datacite.wordpress.com](http://datacite.wordpress.com)

## Dark Data and the Long Tail of Science

Lesley Wyborn, Geoscience Australia & Andrew Treloar, ANDS

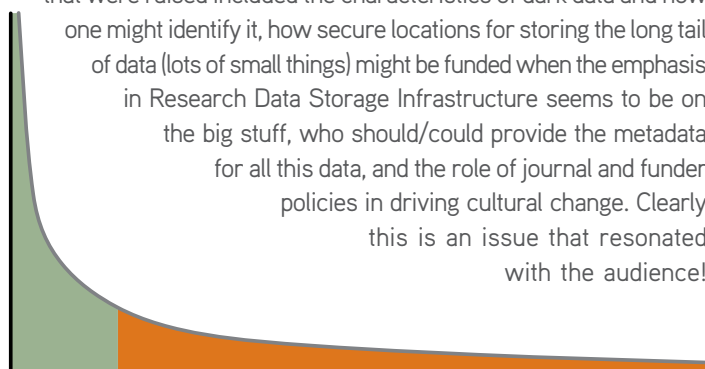
The Dark Data and the Long Tail of Science, Birds of a Feather session at eResearch Australasia 2011 was intended to build on the conference keynote talk by Bryan Heidorn (University of Arizona) on the same topic. More information on the long tail concept visit: [en.wikipedia.org/wiki/Long\\_Tail](http://en.wikipedia.org/wiki/Long_Tail)

Lesley Wyborn, Geoscience Australia ([ga.gov.au](http://ga.gov.au)) set the tone by reminding us why the long tail is important in research. We value the data collected on expensive instruments and fund this well but the software to manage this is low cost and few people are needed. Low cost data (e.g. in situ observations) is far more labour intensive and expensive to store, and needs to be valued when making funding decisions.

Bryan noted that although much research is still field based, we lack good ways to support this. New technologies, like iPhone applications, provide new possibilities to collect data and photos, GPS locations, time, date, etc and allow for transfer to centres for validation and peer review of the data, and then deposit straight into repositories. In this way, so-called citizen scientists can provide large amounts of data for very low cost.

Andrew Treloar, ANDS discussed the ANDS four transformations and their importance for making dark data better managed, well connected and findable in the support of reuse. He noted ways that ANDS-funded data capture projects seek to reduce the costs of capturing data and rich metadata, thus simplifying exposure of dark long-tail data.

The ensuing discussion time was very vigorous and some of the themes that were raised included the characteristics of dark data and how one might identify it, how secure locations for storing the long tail of data (lots of small things) might be funded when the emphasis in Research Data Storage Infrastructure seems to be on the big stuff, who should/could provide the metadata for all this data, and the role of journal and funder policies in driving cultural change. Clearly this is an issue that resonated with the audience!



The Long Tail diagram by Hay Kranen ([haykranen.nl](http://haykranen.nl))

# Reuse and the human factor

Simon Pockley, ANDS

One of the significant themes at eResearch Australasia 2011 was the changing role of the Developer.

Data is only useful if you can do something with it and, as keynote speaker, Prof Andy Pitman, Co-Director, Climate Change Research Centre, UNSW ([www.ccrcc.unsw.edu.au](http://www.ccrcc.unsw.edu.au)) demonstrated; enormous exascale datasets don't just present computational challenges but raise concerns about where the skilled people needed to manage data are going to come from. Small science projects make up the long-tail of eResearch. They often present more complicated data sources and a pressing need for specialist skills.

Paul Walk, conference speaker and Deputy Director of UKOLN ([ukoln.ac.uk](http://ukoln.ac.uk)), warned of the consequences of paying too little attention to that often overlooked species in the information ecology—the Developer.

Developers are often perceived as interchangeable resources. In many universities developers find themselves accepting a series of short-term contracts with few career paths that might lead to strategic decision-making positions. The rising trend towards outsourcing software development is often accompanied by a loss of the internal capacity needed to reap the full benefits of a remote service. Local developers play a vital role in tailoring remotely

delivered software solutions so that they can be fit-for-purpose. Without the local capacity for making critical assessments of the software solutions proposed by vendors, some universities have had cause to regret their choices.

One of Paul's successful initiatives is to encourage a community to form around shared activities through a group known as Developer Community Supporting Innovation ([devcsi.ukoln.ac.uk](http://devcsi.ukoln.ac.uk)). Individual developers are networked with their peers in order to bridge isolation, build community, explore collective problem solving and reuse and share tools and code. The Annual Dev8D event not only delivers cost effective training through peer-to-peer sessions and challenges but also acts to break down the traditional distinctions between developer and researcher.

At eResearch Australasia 2011 there was the sense that a new kind of developer is emerging. This new kind of person has the unique and holistic range of skills needed for the conduct of Data Science. Skills are often discipline specific and range from traditional computer science and mathematics—to communication and art. While Google, Amazon, Facebook, and LinkedIn have made their datastreams the core of their success, more attention needs to be paid by the eResearch Community to the human factors that can nurture these new skills into career paths.

## Chair's report – Ron Sandland

Ultimately the purpose of ANDS is all about the creation and development of research infrastructure, infrastructure that enables the Australian research community to meet its ambitious goals more easily and effectively. The reusability of data is at the core of ANDS' mission.

The four transformations described on the ANDS website, [ands.org.au](http://ands.org.au), relate to enabling the transformation of data that are unmanaged to managed, disconnected to connected, invisible to findable and single-use to reusable. These transformations are the cornerstones of enabling Australian researchers to easily publish, discover, access, and use research data.

Data themselves are increasingly being recognised as extremely important infrastructure in their own right. It is clear from my brief attendance at the eResearch Australia conference, to chair a forum, that the level of understanding of the role of data is being increasingly taken as a given. But the development of this realisation has taken quite a while. As a result, early in the life of ANDS, the immaturity of understanding of the whole data enterprise made it difficult to establish a project to develop metadata stores.

In order to realise the ANDS mission metadata stores must become a vital resource for researchers. Our understanding of the concept

of metadata is also growing to embrace new levels and descriptors of data which might include not just the time of creation of the data and broad aspects of its creation, but also how the data were created, machine settings and programming languages involved in creating simulated data, models, and experimental design details.

I am delighted that the increasing sophistication of understanding of the role of data has enabled ANDS to establish a Metadata Stores project that will contribute strongly to the ability of Australia's researchers to reuse data.

Building on the successful Data Capture projects, offers of up to \$225,000 have been made to each of the institutions involved in that program, the institutions ranked highest nationally in terms of research output. Overwhelmingly the response has been one of strong interest recognising that it is just the right time to address the metadata question. And once again, the institutional focus is proving to be the correct one as it provides the basis for recognition of research data as a key institutional asset.

The results of this project will, I am sure, make a major contribution to the value of the infrastructure being created by ANDS.

## Conference reports

### OAIC Information Policy Conference Report

Cynthia Love, ANDS

The launch of the issues paper: *Understanding the value of public sector information in Australia* was a feature of this recent conference, held on 15 November 2011 in Canberra. Around this theme a number of agency leaders presented on what activities they were undertaking and the challenges they were addressing. The presentations related to all information not exclusively data and in fact, tended to focus on publications rather than data. However, key messages outlined by the Australian Information Commissioner, Prof John Macmillan, were equally applicable across data as well as other forms of information:

- » All government decisions, policies and choices are rooted in information
- » Responsible, comprehensive, integrated information management is a core agency function
- » Government information is a national resource that should be available for community access and use
- » Open government is entwined with the pursuit of democratic accountability, integrity, innovation, civic engagement and customer service

A message in the presentation by DIISRTE Deputy Secretary, Ms Patricia Kelly, that resonated was that evidence based policy development is not always supported by organised access to data and the establishment of a research infrastructure addresses this. Ross Wilkinson, ANDS' Executive Director, took the specific message about data further and presented the ANDS perspective highlighting the value that accessible data adds to the quality of research. He also underlined the fact that the public sector is a rich source of valuable data that enables Australian research.

ANDS is available to provide advice and assistance as public sector agencies work to achieve compliance with the requirement to make their data available.

*Issues Paper 2: Understanding the value of public sector information in Australia* can be viewed at: [oaic.gov.au/publications/papers/issues\\_paper2\\_understanding\\_value\\_public\\_sector\\_information\\_in\\_australia.html](http://oaic.gov.au/publications/papers/issues_paper2_understanding_value_public_sector_information_in_australia.html)

For more information about the conference visit: [oaic.gov.au/news/events/2011\\_conference\\_main.html](http://oaic.gov.au/news/events/2011_conference_main.html)

---

### spatial@gov Conference 2011 Report

Greg Laughlin, ANDS

Another November conference was the spatial@gov conference ([cebit.com.au/2011/conferences/spatial-at-gov](http://cebit.com.au/2011/conferences/spatial-at-gov)) in Canberra. Over 500 delegates from industry, government and private enterprise were brought together to address the following themes:

- » Access to Public Sector Information: the most comprehensively covered theme discussing how spatial communities are dealing with government reforms that increase the accessibility of their information.
- » The Spatial Framework for Virtual Australia: showcased Australian Geographic Information Systems (GIS) and spatial infrastructure—focus on cloud applications
- » Government Spatial Activities in Other Countries: addressing issues concerned with Open Geospatial Consortium ([opengeospatial.org](http://opengeospatial.org)) standards and developments in remote sensing
- » Emergency Response: discussed how location based capabilities were applied in recent emergencies

- » Water Security: focused on data sharing and reuse as key underlying enablers to the Federal Government's Water for the Future plan
- » spatial@defence: discussed the issue of open data to support defence against internal threats

ANDS presented a paper about an ANDS-funded project: The Gazetteer of Australia ([www.mymaps.gov.au/gazetteer](http://www.mymaps.gov.au/gazetteer)). It demonstrated that, with this new piece of infrastructure new links are easily created, within and across research disciplines. It means that anyone can now view a location in a rich environment and can use this for subsequent searches. So, for example, someone interested in old gold mining techniques could use a location, like Kiandra, and use the system to interrogate mining and mineral databases around Kiandra. The linking key is place name and the Gazetteer has over 300,000 of them. The paper generated discussion within the attendee group which augers well for future uptake of the service.



## Building a Culture of Data Citation Workshop

Stuart Hungerford, ANDS

Following his thought provoking closing keynote at eResearch Australasia 2011, Jan Brase opened the ANDS-sponsored workshop, *Building a Culture of Data Citation*. Jan gave a persuasive presentation on the changing requirements of modern research projects for discovering, reusing and citing research datasets, which also highlighted the DataCite activities designed to support that goal. Participants took the opportunity to discuss technical, policy and usage issues surrounding the emerging world of data citation in the research sector.

The workshop was also the occasion for the public launch of the ANDS *Cite My Data*—data citation service. It is now in full production after being refined by some ANDS partners during a pilot phase. This service is built on the international DOI (Digital Object Identifier) and DataCite infrastructure and governance structures.

Data is increasingly being recognised as a primary research output and not just a by-product of research. Data citation plays an important role in this change as it facilitates easier verification and reuse of data, and it can track and record data reuse in the same way as publication citation does. Data citation is therefore another form of recognition for scholarly effort, which researchers can use to measure the outcomes of their research.

If you would like to learn more about services for data citation please contact [services@ands.org.au](mailto:services@ands.org.au) or visit [ands.org.au/cite-data](http://ands.org.au/cite-data)

Information about the ANDS Cite My Data Service is available at: [ands.org.au/services/cite-my-data.html](http://ands.org.au/services/cite-my-data.html)

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

### Capture and Publication of Data on the History of Adoption

Leesa Clausen Brown, ANDS

The History of Adoption research, conducted by Monash University's Prof Marian Quartly and a team of researchers from across Australia, produced a data collection that holds great significance and value to researchers, across various research disciplines, and the Australian public.



Image courtesy of the History of Adoption Project, Monash University

A central feature of this project was the collection of life stories—submitted as either a text or sound file—provided by participants who have had experience with the adoption process.

The research team identified the need for the development of a new data capture process, as existing processes were not efficient or effective. A large amount of data collected in this project may never be able to be captured or reproduced again therefore it was imperative that all the data captured was effectively managed.

This ANDS-funded data capture project developed a streamlined and automated data capture process for both the data and metadata, to enable the valuable data captured in the History of Adoption research project, to be better managed and connected, so that it can be easily discovered and reused in the future.

Having the History of Adoption stories available allows other researchers to reuse the stories in new research both in studies of adoption as well as other kinds of cross-disciplinary research.

In the course of the project key achievements included the development of capture and storage processes for the stories, and associated metadata and a process for generating a story web page with attached story transcript, metadata files and search tags, and the publication of the stories to ARROW, the Monash Library public repository.

This project spawned an eResearch ethics forum which meets regularly, and includes membership from within and beyond Monash University. For more information on the project visit: [arts.monash.edu.au/historyofadoption](http://arts.monash.edu.au/historyofadoption)

## Australian Breast Cancer Tissue Bank

Ingrid Mason, ANDS

Moving from manual to automated processes may seem like a small step, but for the breast cancer researchers at Westmead Millennium Institute, in Sydney and users worldwide of the Australian Breast Cancer Tissue Bank ([abctb.org.au/abctbNew2](http://abctb.org.au/abctbNew2)), the new software developed with ANDS funding affords them valuable time to focus on the microscopic images of cancer tissue and blood samples and their research.

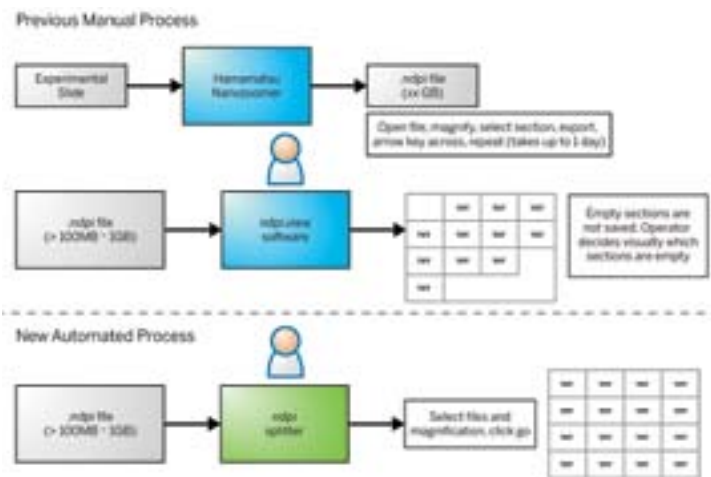
The data management challenges associated with digital microscopy image analysis, a part of observational cell biology research, is being handled more effectively with the new software designed and developed by Intersect Australia. The software captures metadata, splits the large image files into tiles, in a standardized format and at a size better suited for processing and image analysis. Thumbnails are also generated to enable easier discovery using the Tissue Bank search engine.

Prof Christine Clark from the Breast Cancer Research Group at the Westmead Millennium Institute describes the gains made with this new software:

"...data processing which previously would have required days or weeks of laborious image file capture and cataloguing can now be completed in minutes"

The efficiency gains made for the researchers with new software are clear and the potential for research data discovery and reuse have also increased. Rich descriptions of the types of bio-specimens in the Tissue Bank have also been generated to enable researchers that unaware of the research data made available through the Tissue Bank website to find out about this via Research Data Australia.

For more information on the project visit: [intersect.org.au/breast-cancer-microscopy](http://intersect.org.au/breast-cancer-microscopy)



Breast Cancer Tissue Bank process diagram - Copyright © 2010 Intersect Australia Limited

## NeAT's Phenomics Ontology Driven Database Project

Mark Crowe, The Plant Accelerator

Phenomics is the high-throughput study of phenotypes—the physical features of organisms caused by their combination of genetics and environment. Just as with other 'omics technologies, phenomics is constantly encountering new challenges in data management. Projects are highly diverse and generate extensive imaging and metadata, and a way was needed to capture, manage, secure, distribute and publish the resulting datasets.



Fluorescence analysis at the Australian Plant Phenomics Facility

To meet these needs, Australia's two NCRIS-supported phenomics facilities, the Australian Plant Phenomics Facility and the Australian Phenomics Network, recently collaborated on a two year National eResearch Architecture



Infrared analysis of Brachypodium at the Australian Plant Phenomics Facility

Taskforce (NeAT) project with the University of Queensland eResearch Laboratory to develop PODD: the Phenomics Ontology Driven Database ([podd.plantphenomics.org.au](http://podd.plantphenomics.org.au)).

PODD uses semantic web concepts and languages in an ontology-based schema, and the open nature of these makes PODD very amenable to data reuse and interoperability. Unlike relational databases, it is also flexible and extensible and can model any research domain by replacing the schema.

This flexibility is demonstrated by the two phenomics centres operating different instances of PODD, each applying it to data relevant to their particular research domains. Through PODD, metadata from these centres will be linked in to the Atlas of Living Australia as a publicly accessible reference.

For more information on The Plant Accelerator visit: [plantaccelerator.org.au](http://plantaccelerator.org.au)

# Meet the ANDS Staff

## Jeff Christiansen



Jeff Christiansen is a Senior Business Analyst working on the Applications program and various other projects in the biomedical arena. He joined ANDS in April 2009 and works from the Melbourne office.

Jeff obtained his PhD in Biochemistry in 1996 from the University of Queensland (specialising in embryology).

After a post-doctoral research position in embryonic brain development at the National Institute for Medical Research (NIMR) in London, UK, in 2001 he moved to the MRC Human Genetics Unit in Edinburgh, UK to become the first Senior Editor of *EMAGE* ([emouseatlas.org/emage](http://emouseatlas.org/emage)), an online resource describing where and when genes are expressed (or 'turned on') in developing mouse embryos.

Whilst at *EMAGE*, he led a multi-disciplinary team of biologists, image analysts and software developers to take *EMAGE* from an

in-house prototype to a fully fledged public resource. His areas of responsibility covered: securing data from the international research and publishing communities; development and implementation of data management and curation pipelines; quality control; information standardisation; interface and schema design; development of new concepts for data analysis; outreach to users; and working with other staff within the international bio-curation community to ensure consistency of information across resources.

Since returning to Australia in 2009 to join ANDS, Jeff has been providing support to our partners in numerous biology-related ANDS-funded projects, such as the European Molecular Biology Laboratory Australia mirror of the European Bioinformatics Institute. He has also recently started work for a number of projects in the Applications program that span across both various disciplines within the biology/imaging arena, and also a number of partner research organisations. The next few months will see the first of the Applications engagements underway and Jeff is looking forward to working with the many partners involved to help deliver these projects, which are intended to demonstrate the value of having better managed and integrated digital research data, such that novel (and not previously possible) research outcomes can be realised.

## Andrew White



Andrew White's primary role within the ANDS is to provide outreach support for ANDS-funded activities and interests in Queensland. Based in Brisbane, he works across a variety of the ANDS Programs providing project guidance and assistance as well as facilitating outreach activities across a number of Queensland universities and public sector institutions.

Andrew has qualifications in IT and a Graduate Diploma in Library and Information Management. He has worked both nationally and internationally on a range of software development projects as well as with digital repositories focusing on digital collection management.

In his work with ANDS, he finds it rewarding to see the enthusiasm and vision of our partners in helping to build Australia's research data infrastructure despite the obstacles and challenges it poses. Being a part of that and observing it, lead on to a transformation within and outside of the institutions as projects take shape is equally fulfilling, he says.

In the coming months one of Andrew's foci will be working with our partners to examine the process of improving metadata record quality. He will engage with them to develop richly connected exemplar records utilising a variety of tools and standards developed by ANDS. He is looking forward to seeing the benefits this may lead to improving research data discovery and reuse.

## In brief

### West Australia's Australian of the Year 2012

Prof Donna Cross of Edith Cowan University (ECU) has been named WA Australian of the Year for her work in improving the health and wellbeing of youth. She is the Foundation Professor at ECU's Child Health Promotion Research Centre and her research focuses on bullying, nutrition and mental health.

Prof Cross participated on ECU's ANDS-funded Seeding the Commons Data Management Plan and Policy project, which reviewed and redeveloped ECU's data management policy, and increased awareness of research data management within the ECU research community.

For more information: [ecu.edu.au/research/news/2011/11/professor-donna-cross-named-wa-australian-of-the-year](http://ecu.edu.au/research/news/2011/11/professor-donna-cross-named-wa-australian-of-the-year)

### Europe leads the way on reuse of public sector data

The European Commission put forward its Open Data Strategy proposal last December, which is designed to make public sector data more freely available for commercial or non-commercial use, as is already being done in France and the UK.

Neelie Kroes, commission vice president, claims that Public Sector Data is worth more if it is given away. "Taxpayers have already paid for this information, the least we can do is give it back to those who want to use it in new ways that help people and create jobs and growth."

The Strategy also includes the creation of a web-based portal, reported to be in beta phase, which will be the one-stop-shop for reusable data from EU institutions. There will be restrictions on some data, in particular that which is covered by third party copyright.

For more information: [guardian.co.uk/government-computing-network/2011/dec/12/eu-open-data-strategy-neelie-kros-european-commission?newsfeed=true](http://guardian.co.uk/government-computing-network/2011/dec/12/eu-open-data-strategy-neelie-kros-european-commission?newsfeed=true)

### ANDS Gold!

The New Year has seen the commencement of the Research Data Australia (RDA) "Gold Standard" Exemplar Records Projects. These Projects are being run through Queensland University of Technology (QUT) and Griffith University, both early completers of their Seeding the Commons Projects. These projects are the next logical step in the progression of the initial work completed in the earlier projects. The goals of the projects include developing good quality richly connected exemplar RDA records while exploring the process and sustainability of improving the quality of records.

What exactly is a good quality, richly connected RDA record?

It is a record that not only has quality descriptive elements (to improve discovery) but also connects the record to information about the person or group who created the data and where the data was

created; research or project details, the how and why was the data created; and information about the service instrument or tool used to create it. Connecting these elements allows for better discoverability and encourages reuse of research data.

There are a number of tools that these projects will be utilizing to create the Gold Standard records, including but not limited to: Digital Object Identifiers (DOIs), National Library Party Identifiers, spatial information and temporal information.

These projects are designed to provide examples and guidance on Research Data Australia records to other institutions and RDA contributors.

Follow the progress and contribute to these projects via the project blogs:

QUT: [library.qut.edu.au/blog/andsgold](http://library.qut.edu.au/blog/andsgold)

Griffith University: [ands-gold-griffith.blogspot.com](http://ands-gold-griffith.blogspot.com)

Get notified about our forthcoming newsletters via RSS feed: [ands.org.au/newsletter](http://ands.org.au/newsletter)

For more news, alerts, announcements and discussion subscribe to the ANDS General group: [groups.google.com.au/group/ands-general/subscribe](http://groups.google.com.au/group/ands-general/subscribe) or join our conversation on Twitter: [@andsdata](https://twitter.com/andsdata)

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](http://akdesign.com.au))

ANDS Project Partners:



AUSTRALIAN NATIONAL DATA SERVICE

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



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 2.5 Australia License [creativecommons.org/licenses/by-nc/2.5/au](http://creativecommons.org/licenses/by-nc/2.5/au)