

Chair's report - Ron Sandland



Implementation of ANDS' rigorous program is well underway and I am extremely impressed by the commitment of ANDS staff to achieving their daunting goals.

I have recently had several opportunities to see how ANDS can potentially add value to extremely important research in science and society. The first of these was visiting Professor Andy Pitman's Climate Change Research Lab at the University of New South Wales. Climate change data has recently come under considerable scrutiny because it has implications for economic and social policy that are challenging for politicians and the public alike. Much of the data of interest results from running complex models and the provenance of these modelling runs is extraordinarily complex. How can these details be captured at the time the runs are undertaken? It would be well nigh impossible to retro-fit these details after the event. And there are great risks for the researchers in not being able to do so as without this information, model runs may not be able to be replicated and doubts may be cast on their research.

Ross Wilkinson and I visited Andy's Lab with Lindsay Botten, the Director of the National Computing Infrastructure (NCI). There is enormous power to be had in combining the wisdom and resources of ANDS, NCI and ARCS to achieve a whole greater than the sum of the parts.

I am also currently facilitating a process on behalf of DIISR to place an investment in research infrastructure in the built environment. Being able to capture data and appropriately rich metadata, provide open access to data, to define appropriate data management strategies, to publish research data, to deal effectively with data volume and data complexity, to integrate social and physical data. These are just some of the challenges in creating information environments for the research community. But why is this important? Because some of the greatest challenges we face as a nation relate to the sustainability of our urban environments and at present we can pose the questions but do not have the data and models that will enable us to efficiently answer them. ANDS has a crucial role to play (along with our research partners) in providing the underlying technologies that will enable us to do so.

The more I see of ANDS in action, the more I am convinced of the wisdom of the Government's major investment in this area.

Inside Issue 04

- » Chair's report - Ron Sandland
- » Executive Director's report - Ross Wilkinson
- » ANDS out and about - Outreach
- » Conference reports
- » Champions of Research Data
 - » Capturing the reef in real time
 - » History becomes richer through collaboration and data aggregation
 - » Crop research to feed a growing planet
- » Project profile: Australian Research Data Commons Party Infrastructure project
- » Gumboots for the Data Deluge workshops
- » Meet the ANDS Staff - David Groenewegen
- » Research Intensive meeting 19 February 2010
- » Canberra Roadshow
- » Forthcoming events

Executive Director's report – Ross Wilkinson

The success of the Australian Research Data Commons very much depends on the strength of ANDS' partnerships with Government, its collaboration partners, the research institutions, the NCRIS capabilities, the instrument facilities, the eResearch partnerships and suppliers. Most important of all is its partnership with researchers. ANDS staff have been actively engaged in the negotiation of the Data Capture and Seeding the Commons projects at over 30 research institutions. As such the team has been all over the country visiting potential partners and discussing and helping to refine their plans for projects. It has been terrific to see the wide range of proposals and the many exciting new ideas expressed in them. These ideas are rapidly turning from proposals into agreed contracts.

The expectation is that a considerable number of people will be funded and working on activities to capture and manage research data at a large number of institutions around Australia. The majority of these people will be based outside of ANDS. Consequently ANDS is increasingly focused on supporting the development of the collective skill set and knowledge of this broader ANDS community. Please refer to further articles in this issue describing current ANDS activities. We look forward to a rich and vibrant research data community developing across Australia.

ANDS out and about – report on ANDS engagements

Outreach

The ANDS Outreach community is developing. ANDS now has Outreach research analyst capability at eResearch South Australia, iVEC, Intersect and CAIRSS with additional positions planned for other states. Also planned are Outreach positions around specific research endeavours such as Terrestrial Ecosystem Research Network (TERN) and iMOS for the Australian Ocean Data Network (AODN).

The ANDS Outreach community also includes people who are funded for Seeding the Commons activities within research institutions whose focus is primarily on data management and process development within their host institutions. A major challenge in creating an ANDS Outreach community is to find people with a skill set that includes the necessary mix of excellent communication skills and technical awareness.

Conference reports

International Digital Curation Conference, UK eScience All Hands Meeting & IEEE International Conference on eScience

In December 2009, Dr Andrew Treloar (ANDS Director of Technology) visited the UK. Prior to the International Digital Curation Centre Conference in London he presented at workshops on the Citability of Research Data and on Repository Preservation Infrastructure. The actual conference had something relevant to ANDS in every session. It is clear that the DCC has an increasing focus on the importance of data. A particularly good keynote was delivered by Professor D B Kell from the Biotechnology and Biological Sciences

Research Council (BBSRC) – see <http://digitalcuration.blogspot.com/2009/12/idcc-09-keynote-address-douglas-kell.html> for a summary. Dr Treloar presented a paper co-authored with Dr Adrian Burton (Director of Services) entitled 'Designing for Discovery and Re-Use: the "ANDS Data Sharing Verbs" Approach to Service Decomposition' (<http://ijdc.net/index.php/ijdc/article/view/133>). The AllHands conference is roughly equivalent to the eResearch Australia event here. This also had a noticeably stronger data focus than last year. AllHands was followed directly in the same venue by the IEEE eScience conference. Here Dr Treloar presented another Burton/Treloar paper entitled 'Publish My Data: A composition of services from ANDS and ARCS'. Once again, this conference had a strong data flavour to it. It is clear that data is the new black!

VALA Conference

9-11 February 2010

ANDS staff were involved in this conference in a number of ways: as members of the Programme Committee (Treloar), as presenters (Burton/Treloar on 'Publish My Data: the design and implementation of a loosely-coupled data publishing service', David Groenewegen (Director of Research Data) on 'ANDS responses to the data management challenges in the *Australian Code for the Responsible Conduct of Research*'), as session chairs, and as workshop presenters (see separate item on Gumboots for the Data Deluge in this issue of *Share*).

There was an entire day-long track on Repositories, which was actually dominated by papers with a data theme. There was also a presentation by Anna Gerber from the University of Queensland on 'The Aus-e-Lit project: advanced e-research services for scholars of Australian literature'. The main talking points at this conference were around users as co-creators, the tension between assigned and derived metadata and the role of semantic technologies. The conference was also noticeable for heavy use made of Twitter as a backchannel, both amplifying the discussions within the conference and making them available to those who could not attend (Twitter hashtag was #vala2010). Further details, including links to all the papers, at <http://vala.org.au/conf2010.htm>

Champions of Research Data

Capturing the reef in real time

By Pollyanna Sutton (Pollyanna is a freelance journalist)

Australia's Great Barrier Reef is a rich ecosystem and a fertile ground for research. Every day, researchers and environmental groups are generating vast amounts of information in a multitude of forms, from remote sensing satellite pictures to live stream video, numerical data, to maps, and records which look at characteristics including the physical and chemical quality of the water. Until now, building a complete picture of the complex interactions that affect the health and resilience of the reef has relied on trawling academic journals, random sharing with colleagues and the use of USB drives to move data between research labs.

The University of Queensland has initiated the Health-e-Reef project to create a data capture system and sharing service for data and information being collected on the coral reef ecosystems that extend from the Southern Great Barrier Reef (Capricorn Bunker region) to Moreton Bay. The datasets include observational data such as coral cover, coral genetics, algal species distributions, biodiversity data (seagrass cover, fish, stingrays, dugongs, sharks, dolphins, turtles etc), physical and chemical data (turbidity, salinity, sea surface temperature, pH), nutrient data (nitrogen, phosphorus, chlorophyll a, sewage plume mapping), satellite imagery, 3D benthic data, photos, graphs and publications.

In turn Health-e-Reef will provide a unique, comprehensive database and Web interface that enables scientists to understand how the reef might be affected by climate change, urban, industrial and agricultural runoff, fishing and tourism. In order to do this effectively, there is a critical need to share, correlate, and visualise these different types of data. The Director of the University of Queensland Global Change Institute, Professor Ove Hoegh-Guldberg, said that the amount and complexity of the data is growing endlessly. "There are a lot of advantages in being able to share data and see what has been done before because important synergies come from the interrelationships between the data across species, time and space."



A coral community in the marine protected area at Heron Island, southern Great Barrier Reef. Photo courtesy of Prof Ove Hoegh-Guldberg, Centre for Marine Studies, The University of Queensland.

Researchers are currently monitoring how rising sea temperatures may be influencing migration of species – for example, a New South Wales sea urchin recently began invading Tasmanian waters and wreaking havoc on the iconic Tasmanian kelp forests. Professor Hoegh-Guldberg said researchers are looking at factors including sea surface wind speeds, ocean currents and phytoplankton to determine whether they are interrelated in terms of climate change. "How do we make sense of these factors? It helps us understand how to tackle the problem," he said.

Professor Jane Hunter at the University of Queensland e-Research Laboratory has been working on developing the common, extensible metadata schema for describing the coral reef datasets to enable their discovery and re-use. This includes functional specifications for the data set registration, description and publishing services. Semi-automatic metadata extraction will be combined with manual user input to streamline the generation of descriptions of research datasets. Users will also be able to subscribe to RSS/Atom feeds that summarise updates to the dataset descriptions and the addition of any new information.

Professor Hoegh-Guldberg can see the system working on many levels – from secure password protected areas for researchers right down to a public interface where fishermen who might see something unusual are able to upload photographs for researchers to use for tracking trends. He said, "The beauty of the age we live in is that people studying reefs around the world can go to the databases and technologies like the Google Earth platform – which has gone a long way to showing us how we facilitate public access."

There are still many considerations for the Health-e-Reef project including making the interface more user friendly, giving scientists' confidence about the security of their uploaded datasets, and meeting large-scale storage requirements over time. The long term aim is to ensure greater re-use of information and to link the Health-e-Reef database with other related marine datasets information being generated by organisations including the Integrated Marine Observing system (IMOS), and the Great Barrier Reef Marine Park Authority.

History becomes richer through collaboration and data aggregation

By Pollyanna Sutton



Convicts and Diggers: the Claytons, grandfather and grandson, recorded and measured in the Convict records from the Archives Office of Tasmania, and in the AIF records, digitised by the National Archives of Australia. Images courtesy of the Archives Office of Tasmania (<http://search.archives.tas.gov.au>)

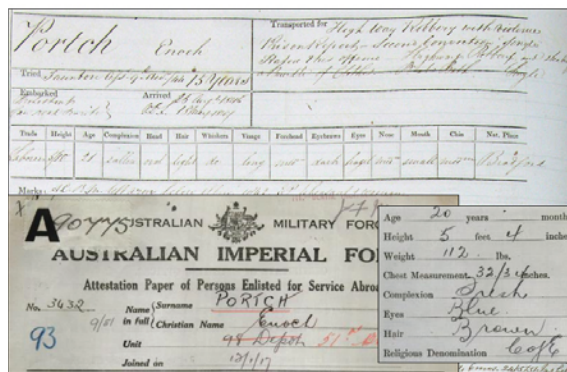
Tracing history once took researchers into archives of dusty volumes, and files of fragile paper. Technology is now giving them access to dynamic datasets that help to build a more extensive picture of who we are and where we come from. The Founders and Survivors project (<http://www.foundersandsurvivors.org/>) is a collaborative effort from researchers, academics and technical staff from the University of Melbourne, University of Tasmania, Australian National University, Flinders University of South Australia, All Souls College University of Oxford and University of New South Wales. It will bring together for the first time diverse information about Australia and its foundations, based on 73,000 Tasmanian convict records.

Participant from the University of Melbourne, Professor Janet McCalman said, "The Tasmanian convict records are considered to be one of the most comprehensive profiles of a people in the 19th century because they track not only the age and physical features, but also health, personality and to some extent the ongoing history of the convicts once they were released." In addition to this Founders and Survivors has gained access to and has created by imaging and transcription, more than 20 datasets about the convicts' lives including births, deaths and marriage records, surgeons' logs on ships to Australia from the National Archive in the UK, links to the Old Bailey website in the UK to track where they may have been convicted, and information about female convicts, ships' records of people leaving the island and military records.

Collecting data began with photographing convict records and gathering more than 100,000 data lines to be converted into prosopographies, or little biographies of people in the context of their times, tracking available information on them from birth to death. This has made the convict records ordered and readable for the first time. The project technology is inspired by the immensely popular Old Bailey website in the UK which allows people to search through all the legal cases that were heard at the courthouse in London from 1674 to 1913. Technical lead for Founders and Survivors, Sandra Silcot, said they are using an XML database, which provides a lot more flexibility in processing arbitrary hierarchies of information.

The underlying technology is an XML encoding standard, TEI (Text encoding Initiative), which was well established in humanities research 20 years ago, and extended about three years ago to deal with prosopographies and social network analysis. "TEI analysis provides ways of dealing with data about people, and the events that happened to them, by using this diverse historical data to create a coherent view of a person – a series of events in a time period that creates a life course narrative," she said.

Professor McCalman said this kind of data will allow researchers to look at some of the social human capital that the convicts brought with them. They can explore – where convicts came from; their physical features and health; exposure to and survival of disease; how height impacts longevity; the effect of mental torment, hard physical labour and a high calorie diet on health. Founders and Survivors like many data projects still faces some restrictions about access to people's information and datasets.



Records for convict Enoch Portch and his namesake, in the First AIF. Image courtesy of Mapping Our Anzacs digitised by the National Archives of Australia: <http://mappingouranzacs.naa.gov.au/default.aspx>.

In this development phase members of the public have been invited to become collaborators and assist with the data. In the future the public search interface will aim to present the best use of materials that are available. Silcot said, "The whole architecture we have adopted can support very in depth research for the research team, while being able to present a public interface to those people who are willing to make their information public. Technically we have been careful to architect that in." "Matching and linking infrastructure has been a very pivotal part of the project and by matching work flow and by deploying this, an online research team can enter any records and have that trigger a live match," she said.

The future plan, which is funding dependent, will allow the database to link with more datasets ranging from the ANZAC records (<http://mappingouranzacs.naa.gov.au/>) to new databases like Plebeian Lives and the Making of Modern London, which will use technical advances in the creation and analysis of digital resources to create a comprehensive electronic edition of primary sources on criminal justice, provision or poor relief and medical care from 1690 to 1800 in London. Professor McCalman said, "This information gives people so much pleasure; they start thinking about both futures and yesterdays."

Crop research to feed a growing planet

By Pollyanna Sutton

In Western Australia the Centre for Comparative Genomics (CCG) at Murdoch University engaged in a project to share information for integrating new data leading to the mapping of the wheat genome along with environmental data to help devise strategies to increase yields on diminishing tracts of land. Professor Rudi Appels and other members of the CCG team are working closely with farmers in Corrigin in Western Australia to gather information to create a web resource that will engage farmers and agronomists to contribute and share critical environmental and soil moisture/nutrition information at the paddock level. A database will enable the aggregation of information for researchers so they can assess how many variables impact on the wheat crop, and build strategies to modify wheat to perform under climate conditions like drought.

There is a pressure on farmers to obtain a good dollar per acre yield and at the same time increase the quality and quantities so they can remain competitive in foreign markets.

Professor Appels said, "We are in effect undertaking the scientific equivalent of the human genome project for wheat, looking at DNA sequences generated with colleagues in China (which has among the most advanced biotechnology facilities in the world). Our project will create an Internet resource where the growers can interact with agronomists and each other to create a common ground for the rapid integration of new and existing data that will generate new knowledge."

The wheat genome project is part of an international collaboration with the International Wheat Genome Sequencing Consortium, the Beijing Genomics Institute, Chinese Academy of Agricultural Sciences, Proteomics International Pty Ltd, the University of Western Australia and University of Adelaide, to study characteristics in an effort to meet the growing demand for food.

"The Federal Government, State Government and growers have been interacting with groups in the Asian region, including China, but when we look around there are huge areas for improvement for the data integration on wheat," he said.



Professor Rudi Appels and a colleague in China examining a copy of an old scroll, 2 metres long, that described the details of village life through images that provided an integration of the jobs people carried out, the crops grown, the equipment used and the structures of facilities. The modern version of the scroll is the internet. Photo courtesy of Prof Rudi Appels, Centre for Comparative Genomics, Murdoch University.

The Australian annual grain output is currently 20 – 24 million tonnes compared to China's 100 million tonnes. Supplying the right type of grain and quality into an already large market place will help Australian farmers maintain a competitive position. This has already been seen with the WA grain which targets quality attributes for the Japanese Udon noodle.

Professor Appels said, "This research will bring together the growers and researching world-wide, and especially nationally, to learn how to make better use of the land and water that is available." Like most datasets there will be access negotiated at a research level and some considerations around security for public access to protect market potential and crop quality implications.

In the next 20 years the human population on the planet will reach an all time high and there will be a conflict around food and water. "The integration of diverse sources of information will provide the basis for better use of available land and water for producing grain targeted to particular markets."

Project profile: Australian Research Data Commons Party Infrastructure project

One of the aspirations behind the Australian Research Data Commons (ARDC) is to link research outputs with the researchers and research organisations who contributed to them and who are responsible for their maintenance and access. Because varying forms of name are used and names for both people and groups can change over time, this is not possible unless there is a common way for systems to reference researchers and research organisations.

A good example is Ove Hoegh-Guldberg, Professor of Marine Studies and Director of the Global Change Institute at the University of Queensland. He has published a number of works over the period of his career with slightly different name forms and different affiliations and has also published datasets in the ARDC. In the case of Professor Hoegh-Guldberg, if all the resource records referenced his unique identifier in the People Australia Service, all of the public resources and datasets he has created could be linked. These could also be linked to information about his blogs, his collaborators, about the Global Change Institute and the University of Queensland.

The National Library of Australia currently maintains the People Australia service, which persistently identifies both people and organisations who have a public identity, who are the authors of published works or who are the subject of such works. ANDS has funded a project to extend the People Australia infrastructure to include Australian researchers and research organisations (parties in the Australian Research Data Commons).

The assignment of identifiers to researchers' public personas will enable a more coherent view of research data collections in the ARDC and their relationship to other parties in the innovation system. It will also enable acknowledgement of those who contribute to the data commons and assist with reporting.

This infrastructure will have many flow-on benefits for the research sector. For example, research administrators, repository managers and data facility managers will be able to uniquely reference the "authors" and producers of the research materials. This new infrastructure will enable many initiatives where information about research and researchers is linked across institutional boundaries.

The ARDC Party Infrastructure Project is well underway and the initial phase includes community consultation on research sector requirements for the identification and description of researchers and research groups. The Australian Access Federation (AAF), Australian Research Council (ARC), and the National Health and Medical Research Council (NHMRC) are being consulted on identifier equivalence services. This project has been receiving valuable support from its Advisory Group comprised of representatives from across the research sector.

There is a clear relationship between this project and other ANDS projects that document the links between research data collections and researchers/organisations.

Keep up to date with news and further information about the project via the ARDCPIP wiki at: <http://wiki.nla.gov.au/display/ardcpip>

Gumboots for the Data Deluge workshops – Melbourne and Sydney

The first workshop in the Gumboots for the Data Deluge series was held in Melbourne on February 12, its timing designed to coincide with the VALA Conference earlier that week. The workshop was narrowly focussed, being designed to assist the increasing numbers of people who need to know more about how to provide records to Research Data Australia. This process starts with data description in the home institution.

The forty participants came from across Australia and as far afield as New Zealand to learn about such topics as permanent identifiers, ANDS services and ISO 2146. Hands-on experience with defining collections provided an opportunity to become familiar with the nitty-gritty of Research Data Australia.

The workshop was offered again in Sydney on April 21. See the ANDS website for forthcoming event details.



Participants at work at the Gumboots for Deluge Workshop held on February 12 in Melbourne

Meet the ANDS Staff – David Groenewegen

David Groenewegen is the Director Research Data and Operations for ANDS, and is based at the Melbourne office. In this role he is responsible for both the "Seeding the Commons" program, which is designed to help increase the amount of research data in the Research Data Commons; and the general operations of ANDS including areas like finance and contracting.

Like many of his colleagues at ANDS, David came to his current role through an unlikely and somewhat roundabout path. His Masters degree thesis at the University of Melbourne was a history of Australian rock music. From there he moved into librarianship, and spent a number of years working in the field at Monash University, mainly in the areas of electronic information provision and information literacy. After a brief secondment to the University of Ballarat he was appointed to the role of Project Manager at Australian Research Repositories Online to the World (ARROW). He went on to combine this with the role of Project Director at the Australian ResearCH Enabling environment project (ARCHER). When those projects wrapped up at the end of 2008 he moved on to ANDS.

David sees the scale of what ANDS is trying to do as the key challenge going forward – it is very much uncharted territory. He notes that, "As far as we know, there is no comparable project in terms of size or funding anywhere else in the world – so we have a huge opportunity to really make a difference." He is looking forward to playing his part in making that difference.



Away from work, David continues to have a keen interest in the history of many forms of popular culture, including cinema, comedy and music. He brings a little of these together with his ringtone – the theme from Groucho Marx's radio show, *You Bet Your Life*.

"...we have a huge opportunity to really make a difference."

Research Intensive meeting 19 February 2010

On February 19, ANDS hosted a meeting at the University of Sydney for representatives of CSIRO and some of the most research intensive universities in Australia. On the agenda was a discussion around ANDS' Seeding the Commons projects, with the aim of the day to facilitate the creation of collaborative projects between Research Intensive Institutions on ANDS Seeding the Commons type activities. It was intended that this would allow us to assess the commonalities of approach and objectives enunciated in the November 2009 EOI proposals to ANDS on Seeding the Commons; and to achieve cost effective and time efficient seeding of the Australian Research Data Commons through collaborative efforts between major Australian research intensive institutions.

The potential collaborative areas were in development of policy and procedure around research data management, particularly in response to the *Code for the Responsible Conduct of Research*; development of training modules and programs around research data management audit of university research data, and development of appropriate data and metadata storage solutions. The meeting was well attended, and some new collaborations were formed. In addition ANDS staff received valuable feedback on where best to focus our efforts in the near future. We are grateful to all those who attended for their time and participation.

Canberra Roadshow



L to R: Martin Borchert (Queensland University of Technology), Andrew Brown (James Cook University), Miriam Saunders (National Centre for Vocational Education Research) at University House for the Canberra Roadshow



L to R: Terri Landford (University of Canberra), Jeff Kelley (ABIN) and Michael James (CSIRO) at the Canberra Roadshow

The first of the 2010 ANDS Roadshows was held at University House at the Australian National University on March 30 and 31. About 30 people attended one or both of the workshop sessions offered. Martin Borchert, who is Associate Director of Library Services (Information Resources and Research Support Services) at the Queensland University of Technology, presented a case study of eResearch Support at QUT as part of the workshop on the *Code for the Responsible Conduct of Research*.

This kind of presentation complements and extends other information and activities offered. There was considerable interest shown in both workshops with high levels of participation and questions.

Roadshows will be held in Sydney on April 19 and 20, and in Melbourne on May 10 and 11. Registration details are available on the ANDS website at <http://www.ands.org.au/events/index.html>

Forthcoming events

ANDS Roadshows

Melbourne, 10-11 May, 2010

Melbourne Town Hall, Swanston Street, Melbourne

Registration details at <http://www.ands.org.au/events/melbourne10-05-2010/registration-melbourne.html>

ANDS Data Capture Briefings

(for those engaged in ANDS data capture projects)

All capital cities and Townsville during June and July

ANDS Data Commons Boot Camp

(by invitation)

Canberra, 7-11 June and 9-13 August

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 Project Partners:



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

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