

Chair's report - Ron Sandland



In my last report I wrote of the challenges of ANDS meeting its 'daunting goals'. While I was, and remain, extremely impressed with the commitment of ANDS' staff to doing what was necessary to meet these goals, I was also concerned about whether the timelines we had given ourselves in the Business Plan were realistic.

So when the opportunity arose to ask the Department of Innovation, Industry, Science and Research (DIISR) whether a short extension might be possible to 'humanise' our deadlines, we were somewhat surprised that there was an opportunity to extend for two years to harmonise with other NCRIS and EIF investments. Because of likely constraints on federal budgets in the immediate future, there would be no new funds to support this extension.

The Steering Committee of ANDS wants to see the creation of enduring value for the Australian research community. So the choice of continuing the work program at the present rate and allowing ANDS to come to a premature end without any chance of continuing funding was not an attractive prospect. Under Ross Wilkinson's leadership we began to explore what an extended ANDS would look like. While our intuition suggested that this extension

would lead to low levels of activity over extended periods of time, this proved not to be the case and, with suitable shifts of funding and timing across reasonably permeable boundaries, a picture of a viable ANDS, continuing to deliver on behalf of the Australian research community, emerged. We realised this was the best way to meet our requirement of an enduring legacy.

Once this decision had been taken it was necessary to negotiate with our partners and DIISR to give effect to the necessary changes in the Business Plan (still underway) and to seek approval from the ANDS Project Partners (Monash University, ANU and CSIRO) for this revision. From my perspective it's a pleasure to deal with the ANDS partners – they ask the right (often difficult) questions but always from the perspective of how can we ensure ANDS' success. Likewise with DIISR who are always willing to help us address difficulties that arise (provided doing so is within their gift of course!). Ross has undertaken extensive consultation with ANDS participants to canvass any concerns about the extension. One theme emerged from these consultations: where deadlines already exist, participants are generally keen to stick to them. This is very healthy: it will promote strong project management and our participants want the outcomes! The continuing support of our stakeholders for the directions we are taking and in helping us to achieve the outcomes is gratifying. Thank you!

Inside Issue 05

- » Chair's report – Ron Sandland
- » Executive Director's report – Ross Wilkinson
- » NeAT Project update on Aus-e-Lit
- » Discovering Australia's Research Data – International Digital Libraries Conference 21-25 June
- » Champions of Research Data
- » Unique museum promotes an improved understanding of architecture
- » Meet the ANDS Staff – Andrew Treloar
- » In brief
- » ANDS Data Commons Boot Camp 7-11 June 2010
- » Forthcoming events

Executive Director's report – Ross Wilkinson

Around the country ANDS-funded research data capture activity is now taking place. Research organisations are deepening their capacity to manage this data. More research data is being described in Research Data Australia. More data is being made visible – research data created locally, national collections, and international collections such as the European Bioinformatics Institute data.

These activities will enable researchers to create more compelling evidence of their research, provide a new form of publication of research and enable improved mechanisms of data citation. Most importantly Australia is creating research data collections that enable more complex data analysis, new discoveries, and exploration across traditional data and discipline boundaries.

This will of course all take time, but it is terrific to see research groups, eResearch organisations and research institutions engaging so enthusiastically.

As can be seen in our reports in this issue of *Share*, many researchers are actively taking up these opportunities to improve how they work with data. ANDS, along with our other eResearch partners, is endeavouring to help provide Australian researchers with a real advantage in research data. In recent interactions with related organisations around the world, we can see that our efforts are consistent with other approaches. But Australia's nationally-coordinated approach has real advantages for its researchers, and the path we are taking is attracting considerable interest around the world.

NeAT Project update on Aus-e-Lit

The Aus-e-lit project will provide several services to support the Australian Literature community:

- » Data integration and federated search
- » Empirical reporting
- » Compound object authoring, editing and publishing
- » Collaborative annotation

So far the project has developed the annotation and authoring tool called "LORE". The SURF and DANS organisations in the Netherlands had expressed interest in examining the LORE tool when it became available as open source. The open source version has now been released.

The project is currently focussed on:

- » refining the LORE user interface
- » integrating federated searches and RSS update feeds
- » visualisation of annotations through tagging

The project is on track to complete this coming financial year. The project is now looking to involve the community in further usability testing and will be promoting the tool at upcoming events to encourage uptake within the community.

More information available at the following links:

Aus-e-Lit project

<http://www.itee.uq.edu.au/~eresearch/projects/aus-e-lit/>

LORE

<http://www.itee.uq.edu.au/~eresearch/projects/aus-e-lit/lore.php>

SURF

<http://www.surf.nl/en/Pages/home.aspx>

DANS

<http://www.dans.knaw.nl/en>

LORE open source tool

<http://sourceforge.net/projects/metadata-net/files/>

"Australia is creating research data collections that enable more complex data analysis, new discoveries, and exploration across traditional data and discipline boundaries."

Discovering Australia's Research Data – International Digital Libraries Conference

21 – 25 June 2010

For the first time, the ACM/IEEE Joint Conference on Digital Libraries (JCDL) and the annual International Conference on Asia-Pacific Digital Libraries (ICADL) were held together. The conferences took place from 21-25 June 2010 at Surfers Paradise in Queensland. Dr Stefanie Kethers, an ANDS Senior Business Analyst presented a JCDL short paper entitled 'Discovering Australia's Research data', co-authored with ANDS colleagues Dr Xiaobin Shen, Dr Andrew Treloar, and Dr Ross Wilkinson. The paper explored how the Australian Research Data Commons can help researchers address

broad, complex, and cross-disciplinary research questions by providing web pages that describe data collections with rich links and rich text.

The conference had great keynotes by Katy Börner (Indiana University), Curtis Wong (Microsoft Research), and David Rosenthal (Stanford University). David's keynote was particularly fascinating from an ANDS perspective; an edited version of his talk can be found at <http://blog.dshr.org/2010/06/jcdl-2010-keynote.html>

More information is available on the Conference Web site, <http://www.jcdl2010.org/>

Champions of Research Data

Sharing oceanographic data with the world

By Pollyanna Sutton (Pollyanna is a freelance journalist)



A female southern elephant seal at Macquarie Island, weighing more than 350 kg, carrying a CTD-SRDL tracking device (SMRU, St-Andrews University). The tracker will record seal diving behaviour and the temperature and conductivity of the water, and send back almost in real-time via the Argos satellite system, while the seal is at sea in the Southern Ocean for up to nine months. Photo courtesy of Chris Oosthuizen.

As elephant seals make their way from Macquarie Island into the Southern Ocean, some will carry miniature tracking devices on their nine-month odyssey, to allow scientists to observe their foraging behaviour and the ocean around them. It is a great example of how two completely different areas of science can use one tracking device to gather valuable data. As biologists at Macquarie University and the University of Tasmania gather data about the behaviour and feeding habits of the seals, collaborating oceanographers at CSIRO will use the measurements of temperature, depth and salinity, to build a picture of the effects of climate change. The project is led by Dr Iain Field, Deputy Director Marine Mammal

Research, at Macquarie University. The data will be used in a collaborative research programme between Professor Rob Harcourt from Macquarie University, and Professor Mark Hindell from the University of Tasmania.

Elephant seals are well studied on land but their ocean behaviour is less documented. Currently, there are 15 seals with devices, which will stay attached until their coats moult in the spring/summer season. Dr Field, said there is no detectable impact on the animals, and they are providing information from locations including Commonwealth Bay in Antarctica. In the past oceanographers have relied on Argo floats to gather data, automated devices that descend into the ocean to a cruising depth and then re-emerge to transmit data to satellites. The seals are able to go to and under the ice, where the Argo floats are unable to travel, thus allowing measurements in a previously unobserved region.

This project is part of IMOS (Integrated Marine Observing System), an Australian Government initiative that is building a sustainable ocean observing system with the collected data available to be shared around the world. Once the seal tag information has been collected it is fed back into the eMarine Information Infrastructure, an IMOS facility at the University of Tasmania. eMII coordinates how the data is handled, stored, accessed and promoted. eMII collaborates closely with both ANDS and ARCS.

Katherine Tattersall is a project officer for IMOS eMII in Hobart. She explained how the data collected from the seals is sent via satellite to the Sea Mammal Research Unit in St Andrews, Scotland, which has an automated system to decrypt it, and automatically place it in an online data repository. eMII then retrieves the data once a day. "From seal to satellite to Scotland and then back to Hobart, sometimes the data is available in Hobart within five hours of the seal collecting it," Tattersall said.

Accessing the IMOS Ocean Portal (<http://imos.aodn.org.au>) allows researchers to see information including the seal tracks relative to the Australian continent, the position of Argo floats, and other geophysical information. The interface is simple, graphic and user friendly. Data is also fed back into the Australian Ocean Data Network (AODN), a shared network for Australia's major marine data agencies, - Australian Institute of Marine Science (AIMS), Australian Antarctic Division (AAD), Bureau of Meteorology (BOM), CSIRO Marine and Atmospheric Research (CMAR), Geoscience Australia (GA) and the Department of Defence (RAN Directorate of Oceanography and Meteorology).

Tattersall said, "There has been a lot of isolated data collection in the past, now we are seeing the benefits of sharing data; it gives everyone access to many types of information that they would not have had before. Data sharing is becoming a global activity, involving groups in Europe, the US and other parts of the world."

www.imos.org.au

<http://www.aodc.gov.au/>

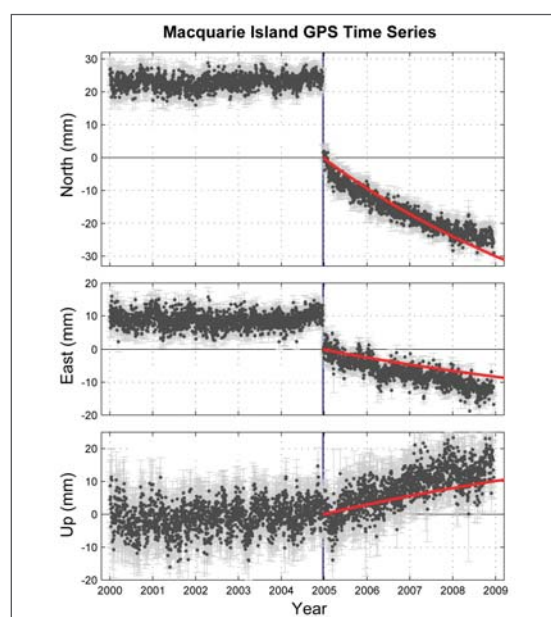
From Mawson to TerraWulf II - Modelling the Australian Continent

By Pollyanna Sutton

Australian scientists are using state of the art satellite technology combined with sea level records collected during Mawson's expedition to Macquarie Island in 1912-1913 to define how the Australian continent is changing. Harvesting data from GPS (Global Positioning System) devices on Macquarie Island and others from around the world helps to build a picture of the Earth's moving plates and the effect of events such as earthquakes on land deformation. Dr Christopher Watson from the University of Tasmania (UTAS), and Dr Paul Tregoning from the Australian National University (ANU) have been running historical GPS data from the last eight or nine years through calculations, using the new TerraWulf II computer facility at the ANU. The ANU super computer's capacity means calculations that once may have taken these researchers over 20 years can be done in a matter of weeks.

A GPS receiver receives signals from satellites orbiting the Earth that are used to calculate the distance from the satellite to the device every 30 seconds. These complicated calculations end up using 24 hours of these observations to provide the latitude, longitude and height of the GPS site on the ground - basically one dot on a chart. The dots give a time series and allow scientists to infer uplift in the land and plate motion. The GPS data can be accessed through the AuScope grid, an infrastructure system being developed to give researchers access to a range of technologies and capabilities in data acquisition, management, modelling and simulation across the geospatial and geoscience spectrum. Dr Tregoning said, "The outcome is that we have been able to reduce the noise and get greater accuracy in position. We can better model the effects caused by tides and reduce the impact of the atmosphere on the signals as they propagate from the satellites."

Macquarie Island sits on the edge of the Australia/Pacific plate boundary, on a ridge known as MRC. This area has experienced many historical earthquakes including two of the largest transform earthquakes in instrumental records, in 1989 and 2004. Analysis of the GPS data including the Macquarie Island station has given them insight into the amount of 3D deformation from regional earthquakes.



The figure shows the changes to the Macquarie Island GPS station position over time. The underlying tectonic plate motion has been removed from the time series, leaving the effects of the large earthquake that occurred on 23 December, 2004. A large jump can be observed at the time of the earthquake, followed by a change in the direction of the motion at the site. The red lines show a model that attempts to predict the changing motion at the site induced by this earthquake. Understanding the response of sites such as this to earthquakes aids in our ability to interpret sea level changes in an absolute sense - this is particularly important in regions such as this where very few sea level data exists. Courtesy of Dr Christopher Watson, University of Tasmania

The movement in the vertical direction is particularly important for sea level studies. Using the tide gauge data from Mawson's expedition has become a strong element in understanding the real measurement of sea level rise, when the land movement is taken into account.

Dr Watson and Dr Tregoning found that Macquarie Island has been moving: as sea levels rise, the land has been subsiding in some areas, giving the impression that sea level rise is higher. Following the 2004 earthquake, the rate of the subsidence slowed down. This was an important clue in better understanding the way the Earth responds to earthquakes. Dr Watson said, "In the slow movement of the earth, stress builds up leading to earthquakes. This can be followed by a long period of relaxation in the crust - we are still seeing the effect of the big earthquake in 1924." AuScope is funding new tools to access high powered computing resources over rich data through web interfaces, and ANDS is funding improved access to spatial data resources.

Dr Tregoning said that once it is effectively rolled out this will allow streamlined access and give students and researchers the chance to run different solutions to see what happens for example with crustal motion.

"The ability to run sophisticated models using high powered computers has revolutionised how we do research", said Dr Tregoning. Dr Watson said, "This is fundamentally changing the type of analysis we do, changing where we look for answers to problems and giving us far greater ability to test for different things at the nitty gritty level."

"The ability to run sophisticated models using high powered computers has revolutionised how we do research"

Collaborating to catalogue cancer mutations

Scientists involved in the International Cancer Genome Consortium (ICGC) (www.icgc.org) have waived their patent rights so that other researchers are free to access data for cancer research. The ICGC is a major collaborative program involving scientists from fourteen countries. "Cancer research can happen in a timely fashion only if researchers collaborate with others to share ideas, resources and data," comments Professor Sean Grimmond.

Currently, two ICGC projects are underway in Australia. Grimmond is the Principal Investigator on the Australian Pancreatic Cancer Genome Initiative (APCGI). The project is carried out in partnership with Professor Andrew Biankin from the Garvan Institute of Medical Research, who has established a national network for sample collection, clinical data acquisition and the clinical review of specimens. They have also partnered with Professor David Bowtell from the Peter MacCallum Cancer Centre for the Australian Ovarian Cancer Study (AKA AOCs). All three investigators have research programs that run in parallel and that are complementary to the ICGC effort.

All ICGC member organisations reached agreement on how they would work together to catalogue the genetic changes of the 50 most common cancers. Through a lengthy consultation process it was agreed that a set of common standards would be used for describing the disease and pathology terminology, and for the format for capturing images and data. An agreed process for sharing the data was also developed.

The Queensland Centre for Medical Genomics (QCMG) (www.qcmg.org) within the Institute of Molecular Bioscience at the University of Queensland carries out all the sequencing, data management, and bioinformatic analysis for these projects. Grimmond said, "What we are seeing now is a technological revolution. A genome can now be sequenced in under a month and for a fraction of the previous cost. We can now work out the genes that are driving the cancer and create a thorough atlas of the DNA.



Research assistant, Senel Idrisoglu uses a gene sequencing machine. Photo courtesy of the Queensland Centre for Medical Genomics, Institute for Molecular Bioscience

"Applying the genetic findings will lead to new treatment solutions for patients." To date they have released the data for four pancreatic cancer genomes and are aiming to sequence the normal and cancer genomes for 375 pancreatic cancer patients and for 125 ovarian cancer patients over the next four years.

All the processed data are available through the portal hosted by the Data Coordination Centre based within the Ontario Institute for Cancer Research. The public can access summaries of the data, and any information that is anonymous and not considered ethically sensitive. Researchers are able to access detailed data by gaining approval from their local ethics committee and the ICGC's ethics committee. Researchers can ask different questions about every sample that has been collected.

Because of the sheer scale of the data being generated it is not possible to store it indefinitely at the Institute of Molecular Bioscience. Currently, the raw data are deposited in the European Bioinformatics Institute (<http://www.ebi.ac.uk/embl/>) data facility. ANDS is one of several organisations providing funding for the development of a mirror of the EBI data facility in Australia.

Continued on page 6

"The EMBL Australia EBI Mirror will help EBI better serve research communities here in Australasia and Asia-Pacific, not only by distributing workload geographically but also by providing customised data services in areas of Australian research strength such as plant biotechnology or biodiversity genomics," said Professor Mark Ragan, Head of Genomics and Computational Biology at the Institute for Molecular Bioscience.

"In time, the Mirror will incorporate unique capabilities and software developed in Australia, and make these available to researchers globally. Existing and new EBI data that originate from Australia will be made available for discovery through the Australian Research Data Commons. They will also be assigned persistent identifiers that allow them to be combined into unique data products that address issues such as climate change or coastal management."

Unique museum promotes an understanding of architecture

In the mid 1970s architectural historian Donald Leslie Johnson began collecting drawings, practice records, personal papers, photos, slides, books, and periodicals of significance to South Australia's architectural history. The collection grew, and in 1990 Leslie donated it to the School of the Built Environment of the South Australian Institute of Technology (SAIT). SAIT is now the University of South Australia (UniSA), and the materials that Leslie collected form the basis of the present day collections of the UniSA Architecture Museum, located at UniSA's CityWest campus in Adelaide. The Architecture Museum's collections now consist of over 200,000 items, including approximately 20,000 drawings, 4000 antique prints, and a 2500 volume library.

The Museum's collections are unique in their scope and focus because they represent a range of private practitioners, rather than only well-known figures or architectural 'heroes.' In addition to their relevance to research in architecture and related fields, the Museum's collections are relevant in the Humanities, Arts and Social Sciences more broadly because they contain data pertinent to many historical, social and cultural themes.

The Architecture Museum's collections are currently utilised by a number of researchers, especially including staff, and postgraduate, and undergraduate research students at UniSA. Many publications have resulted from researchers' work with the Architecture Museum's collections, and the Museum mounts regular exhibitions, sometimes in collaboration with other research and historical bodies.

ANDS is now working with the Architecture Museum to achieve several outcomes. The project will improve the Museum's internal processes around the description of its holdings, and the upload of collection descriptions to a UniSA institutional data repository. In the future, there is potential that the processes developed during the project will be applied with other UniSA research data collections, thus improving the University's research data management practices. Significantly, the project will bring the Museum's collections into the Australian Research Data Commons and, as a result, the Architecture Museum's collections will become visible, and hopefully utilised for research, worldwide.



Dr Christine Garnaut, (Director, Architecture Museum, School of Art, Architecture and Design, University of South Australia) shows Dr Andrew Williams (ANDS) architectural drawings from the Museum's collection.

It is significant that ANDS is funding this work for two important reasons. Firstly, the Architecture Museum is the only such institution in Australia. Secondly, the Architecture Museum is also the only Australian institution or agency member of the International Confederation of Architectural Museums (ICAM), the body dedicated to fostering links between all those interested in promoting an improved understanding of architecture. The Architecture Museum's participation in both ICAM and the Australian Research Data Commons speaks volumes of the Museum's commitment to sharing information and data and to facilitating research internationally. It is anticipated that this project will open up new opportunities for national and international research partnerships based around the Architecture Museum's collections.

Take a look at the Architecture Museum's current presence on the UniSA website: <http://www.unisa.edu.au/artarchitecturedesign/architecturemuseum/default.asp>

Meet the ANDS Staff – Andrew Treloar

As ANDS Director for Technology, Dr Andrew Treloar takes an overall view of technical issues. In addition, he is the Director responsible for the Data Capture and Metadata Stores programs.

He has a usefully diverse discipline background: a B.A. hons. (first class), majoring in Germanic Languages and Linguistics, a Grad. Dip. in computer science, and an M. A. with the topic A Computer-assisted analysis of characterisation in Virginia Woolf's 'The Waves', all from Melbourne University. In 1999 he received his Ph. D. from Monash University with the topic Hypermedia Online Publishing - Transformation of the Scholarly Journal. His research interests include data management, institutional repositories and scholarly communication.

In 2008 he led the project to establish ANDS. Prior to that he was associated with a number of eResearch projects as Director or Technical Architect: ARCHER (<http://archer.edu.au/>) – an eResearch support environment, DART (<http://dart.edu.au>) – data acquisition and analysis, and ARROW (<http://arrow.edu.au/>) – institutional repository software. He also led the development of an Information Management Strategy for Monash University (<http://www.monash.edu.au/staff/information-management/>).



Over the coming months Andrew is looking forward to seeing some of the innovative ANDS-funded data capture projects starting to deliver, and the amount of data in the Australian Research Data Commons continuing to grow in amount and richness. He is also turning his attentions to the issue of improving the linkages between publications and associated digital content (such as datasets and computer models). As well as being directly relevant to ANDS, it also serves to bring together his research concerns from the latter half of each of the last two decades.

"...improving the linkages between publications and associated digital content..."

In brief

National Science Foundation announces change of policy on data management

In May 2010 the US National Science Foundation (NSF) announced a change in the implementation of the existing policy on sharing research data. The NSF is planning to require that all grant applicants submit data management plans. This is part of an effort to ensure open access for data arising from publicly-funded research. Researchers will soon be asked to submit a data management plan as a two-page supplement to any regular grant proposal. This means that it will become an aspect of the merit review process. As science has become more data-intensive and collaborative, researchers from different disciplines now need to work together to solve complex problems. Increasingly, scientific breakthroughs are going to be powered by advanced computing techniques that allow researchers to explore and mine datasets.

JISC MRD Research Progress Workshop

JISC held a workshop on 17-18 May 2010 in Manchester, UK to consider how best to meet the challenge of managing research data in UK Universities. Participants were informed of progress made under the JISC MRD Programme to examine researchers' data management requirements and improve planning in a number of academic and institutional contexts. The objective was to discuss the detail of the projects' discoveries and to consider the more general landscape and initiatives required. The workshop attracted a range of participants representing various stakeholders keen to improve the management of digital research data in UK Universities. An overview of the MRD workshop programme, together with links to the presentations delivered, is available at: <http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmevents/programmemeetingmay.aspx>

ANDS Data Commons Boot Camp

7 – 11 June 2010

In June ANDS facilitated the Data Commons Boot Camp at the Australian National University in Canberra, where the weather was cold but the hospitality was warm, and the information flowed thick and fast for a week. The participants were all involved in research data management and ANDS-funded projects at Australian Universities.

Over the course of the week the participants all received a great deal of information about eResearch, particularly in relation to research data management, and specifics of engaging with ANDS to make research data visible in the Australian Research Data Commons. They heard from an impressive array of presenters, including ANDS staff, experts in several areas at the Australian National University including the ANU Supercomputer Facility, PARADISEC, the Australian Research Collaboration Service, the National Computational Infrastructure, the National Library of Australia, and CSIRO.

Participants' feedback was positive about the value of the Boot Camp initiative, with one participant commenting that it was, "...a great opportunity to build our community."



L to R: Martyn George (University of South Australia), Frankie Stevens (University of Sydney) and Tom Ruthven (University of NSW) work together on a hands-on activity.

"...a great opportunity to build our community."

Forthcoming events

ANDS Data Commons Boot Camp

9 – 13 August 2010

A five day intensive workshop
Australian National University, Canberra.

Details at <http://ands.org.au/events/bootcamp.html>

ANDS Data Capture Briefing and Roadshow

Townsville – To be advised

Dates and registration details forthcoming at

<http://ands.org.au/events/index.html>

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/>