

Champions of Research Data



A Sleepy Lizard with a data logger. Photo courtesy Jana Bradley, Flinders University

Investigating sleepy lizard ecology

Professor Michael Bull has been studying sleepy lizards for over 15 years, and is a world expert on the species. He and his research team from Flinders University examine various aspects of sleepy lizard behavioural ecology, including formation of stable pair associations, competition for mates, social behaviour, aggression stimulus response, and the spread of pathogens and parasites. In one study of the many published by Professor Bull and his team over many years it is observed that, although the Australian sleepy lizard retains long-term pair fidelity for up to 21 years, in some cases pairs separate, and abandoned males have significantly higher tick loads in their last year with their previous female partner than did males that retained their partners from one year to the next. It seems that parasites are not good for long-term sleepy lizard partnerships.

Professor Bull's research team has developed data loggers that are placed on the backs of sleepy lizards at a research location in the Mount Mary region, just over 20 kilometres southeast from Morgan in the South Australian river land. Every two minutes

a data logger records the number of steps a lizard has taken, along with the ambient temperature and light conditions. Every ten minutes a GPS unit records the lizard's location, and a radio facilitates finding of the lizard every two weeks for data download and battery replacement. The data logging system generates about 140,000 data records per lizard over the six months the lizards are active, providing a complete record of activity levels. Data from the lizards allows phenological correlations with data from local weather stations, and insights into the impact of variable climatic conditions.

Professor Bull collaborates with colleagues at the Graduate Group in Ecology at the University of California Davis, and with his postgraduate and post-doctorate students. He sees potential for his data to be useful to other lizard behavioural ecologists, but also to climate researchers locally and all over the world. Over his career, Professor Bull and his colleagues have observed the Mount Mary sleepy lizard ecological system through two droughts and one of the wettest periods on record.

...continued on page 2

Inside Issue 07

- » Champions of Research Data
- » Chair's report – Ron Sandland
- » Executive Director's report – Ross Wilkinson
- » Addressing the pending shortfall in global food production
- » MyMaps.gov.au website to be deployed early this year
- » Project Update on Australian Research Data Commons - Party Infrastructure project
- » ANDS activities at the eResearch Australasia conference 8-12 November 2010
- » Forthcoming events

This data provides a solid baseline against which future climatic change can be observed and analysed. The data that Professor Bull and his team collect about sleepy lizard parasites also provides a baseline for analysis of exotic pathogens, a new "hot" paradigm for study in the biological sciences.

Professor Bull sees open data as a good thing, leading to new and faster discoveries in his field because it enables easier application of new analytic tools to data. Openness must be balanced against researchers' proprietary feelings toward their data, and their instinct to guard it for their own use in order to generate as many publications as possible.

He sees that research data's shift to a new status as a primary research output alongside publications will require a cognitive shift on the part of researchers, because the creative act of data interpretation required in order to write a paper is currently valued more highly than the initial research effort that generates the data.

A late career researcher, Professor Bull sees the ANDS-funded project and the resulting open provision of his data, carefully gathered and curated for over 15 years as a solid legacy, and an opportunity to ensure the data has an ongoing useful life.

Polar Cloud Encourages Information Sharing

By Pollyanna Sutton (Pollyanna is a freelance journalist)

As the International Polar Year (IPY) drew to a close in June 2010, the launch of an organised virtual storage and retrieval space known as the Polar Information Commons (PIC) gave scientists and researchers a new facility for sharing information and keeping pace with the rapid change in Polar systems. PIC is an initiative of IPY to establish a framework for the long term stewardship of polar information and provide a fast and easy-to-use open data resource that is accessible through normal search and browse tools on the internet. To develop the PIC, committee members spent a year designing a system that would allow researchers to upload data into a cloud facility, and users to access this information through a network of centres throughout the world. Established data centres will eventually serve as repositories for the uploaded information and also important data that has been proactively sourced and assessed through national and global data systems. In order to make this international collaboration work, PIC required a set of standard tools, eager participants and protocols that would ensure that critical data was preserved in the long term.

One of the main drivers for PIC is the Scientific Committee for Antarctic Research (SCAR) which was formed in 1958. SCAR has five standing committees, three of which focus on research and two on data, and operates as an umbrella organisation for many scientific collaborations. Manager of the Australian Antarctic Data Centre, and Chief Officer of the SCAR Committee for Antarctic Data Management, Kim Finney, presented the PIC Cloud to the International Polar Year conference in Oslo, Norway last year. Ms Finney said there are currently four components to PIC: a digital badge that will brand data as belonging to the PIC, a set of norms and behaviours expected of PIC data providers and users; a cloud-based data storage facility and a network of PIC data repositories. PIC badges serve to help public search engines identify polar information and ATOM feeds on the PIC Cloud site alert participating archiving centres to new information that is available.

The PIC Cloud was developed and built by the Tasmanian Partnership for Advanced Computing (TPAC) and the Australian Antarctic Division (AAD) in collaboration with the Australian Research Collaboration Service (ARCS) and ANDS. The backbone of the Cloud architecture uses the ARCS Data Fabric and the ANDS Research Data Australia

metadata system and registry. The PIC Cloud facility is initially only about storage and retrieval with a deliberately limited range of capabilities. At the moment users can deposit up to 10 datasets, with a maximum of 20MB per file, with minimal metadata entry to make the information discoverable.



A screenshot of the PIC website. Image courtesy of Australian Antarctic Division

The metadata is then published to the ANDS metadata registry where the system makes data packages available to public search engines. Through this open access, cloud-based process, users are able to discover, share and exchange data and information without necessarily registering or depositing data directly in more formal repositories or catalogues. The PIC will begin by harnessing centres already participating in

the ICSU (International Scientific Council) World Data Centre System and over time will be expanded to include other associated networks of dedicated polar data centres, which will boost the capacity needed to manage, integrate, synthesise and publish polar information.

<http://www.polarcommons.org>



Kim Finney (on right) (Australian Antarctic Division) explains PIC to Prince Albert of Monaco at the International Polar Year Open Science Conference in Oslo. Photo courtesy Dave Connell, Australian Antarctic Division

Tracking the Southern Ocean

By Pollyanna Sutton



Testing components of a mooring at CSIRO, Hobart. The mooring will be one of three deployed from the *Aurora Australis* (behind) this month in the Southern Ocean near the Mertz Glacier. The data will be retrieved in two years. Photo courtesy of CSIRO

As the *Aurora Australis* sails to Antarctica and the *Southern Surveyor* makes its way across the Great Australian Bight, a new initiative will allow scientists and researchers to track not only their journeys, but also oceanic and atmospheric data collected on the way. These two vessels are key elements of Australia's national research infrastructure, enabling oceanographic, geo-scientific fishery and ecosystem research from Australia's coastline to the Antarctic. Data collected on the ships using an array of instruments will help researchers like Bronte Tilbrook at CSIRO Marine and Atmospheric Research track how much carbon is being absorbed into the Southern Ocean. "Oceans are a major carbon sink and they take up to 25 per cent of global emissions each year. The measurements are helping us determine the carbon uptake and how that is changing," he said. Tilbrook explained that the Southern Ocean is known to be one of the most important sink regions on earth for removing carbon dioxide from the atmosphere. The existing measurements are sparse and the work done in Australia will be a major contribution to an international effort to track the Southern Ocean CO₂ sink.

The National Research Vessel Project is funded through Australia's Integrated Marine Observing System (IMOS). Ships are being fitted with the automated instrumentation and data transmission to shore based facilities and onboard staff are used to monitor the systems. By utilising multiple ships that take different routes around Australia, researchers will be able to build a more comprehensive understanding of the marine environment and how it is changing.

The National Research Vessel Project will capture a range of documented and qualified datasets, transmit them via satellite or the 3G network to the relevant organisation where they will be checked for quality. Some data sets will be available in real time, others will require more comprehensive checking.

Once the organisation (Australian Antarctic Division for *Aurora Australis*, and CSIRO Marine and Atmospheric Research for *Southern Surveyor*) has processed the data, the metadata will be harvested by the Australian Ocean Data Network (AODN) and made discoverable via the information system of the AODN portal (<http://www.aodn.org.au/webportal>) and then harvested into the Australian Research Data Commons (ARDC). Director of the IMOS e-Marine Information Infrastructure (eMII) at the University of Tasmania, Dr Roger Proctor said this project has two phases: first the development of common data publication protocols and standards for the data from each vessel, and second, delivering the data and metadata via web services and converting the metadata to RIF-CS, an XML schema used to structure information, enabling the metadata to be shared with Research Data Australia (RDA). The first two vessels will serve as sophisticated collection platforms, capable of capturing data about the oceans, the sea bed and atmospheric conditions. A third research ship, the *Astrolabe*, is also being fitted with instruments.

Tilbrook said that while qualified staff are essential to keeping the instrumentation in good working order, the automated data collection reduces the time researchers have to spend at sea. The near real-time transmission of data also allows problems to be identified rapidly and corrected, maximising the data returns. "The new technology we are using is more cost effective and vastly improves coverage," he said. Aligning information from two different vessels and automating data publication also has many advantages for the global research community. Accelerated access to data will mean faster reporting and modelling of the changes in weather, ocean ecosystems and marine life and in turn the effects of climate change. On the AODN data portal people can see an example of how the interface will work, by clicking on the voyage trail left by the *Aurora Australis*. A popup screen then shows which preliminary data has been collected at that time and position, and indicates other types of data which will be shown once the system is fully implemented.

ANDS was pleased to be able to provide financial support to enable the capture and discovery of this data.

<http://portal.aodn.org.au/webportal/>

"The new technology we are using is more cost effective and vastly improves coverage,"

Chair's report – Ron Sandland

2010 was another complex and challenging year. The highlights were many and, as Chair of the ANDS Steering Committee, it never ceases to amaze and gratify me that every challenge that is thrown at the ANDS team seems to be met with a constructive and appropriate solution. It behoves me on behalf of the Steering Committee to congratulate Ross and his team on an outstanding year.

Some of the challenges met during 2010 included signing up virtually all of Australia's major research institutions to significant data capture projects, working through how to completely change (extend) the timetable over which ANDS will be delivered and convincingly articulating what will change as a result of ANDS' existence. In each of ANDS' programs the results are impressive. We have talked a great deal about Data Capture and Seeding the Commons but significant advances are being made in the ARDC Core Program, for example, where the development of the ANDS portal, Research Data Australia, in which over 1,000 data collections are being made visible and institutions are increasingly alive to preparing records for release - with almost 4,000 records in the Sandbox.

An important event on the ANDS calendar was the December AeRIC meeting which took a strategic overview of the national eResearch capabilities. AeRIC is the strategic committee that has oversight of the eResearch capabilities. Its membership comprises representatives of the various eResearch capabilities (ANDS, NCI, NeCTAR, ARCS ++) as well as eResearchers from a wide variety of disciplines and sectors. The meeting had representatives from industry who made an invaluable contribution – the data tsunami was seen as one of the critical drivers of the eResearch revolution.

Critically, DIISR intends to refresh the roadmapping process which led to significant national investment in eResearch infrastructure. It is clear that this process will underpin crucial decisions on the nature and quantum of future eResearch infrastructure investments. From ANDS' perspective, having articulated our value proposition, we must demonstrate how, where and by how much we are changing the way research is done. I believe this is critical to national productivity and competitiveness but we need to nail this argument.

The national eResearch Australasia Conference was notable for the dominance of data-related issues in presentations. The understanding is building as to just how critical it will be to deal with the challenges of the data tsunami. But the issue for ANDS remains to demonstrate not just the pivotal importance of data in eResearch but also how ANDS is helping the researchers to deal with it productively and efficiently.

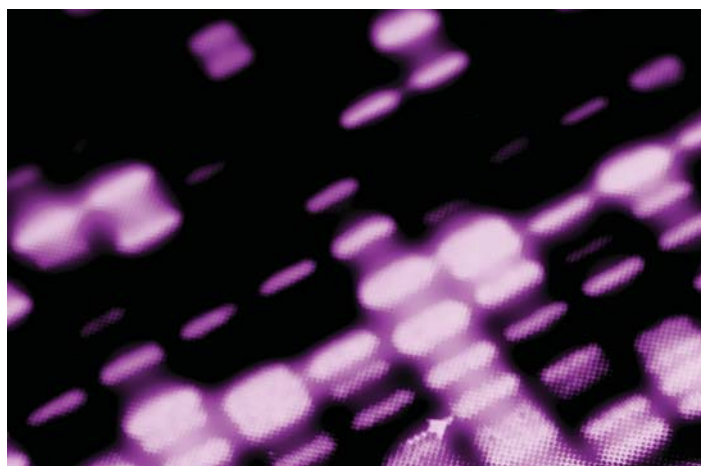
Finally, the ANDS Steering Committee is delighted to welcome two new members from the beginning of 2011: Professor Craig Johnson from the University of Tasmania, an eminent marine ecologist who is passionate about data. Craig will strengthen the research focus of the Steering Committee. And Dr Siu-Ming Tam, First Assistant Statistician, Integrated Collection and Dissemination Services with the Australian Bureau of Statistics, who will significantly enhance our understanding of the issues related to public sector data.

From all of us in ANDS we wish you all a wonderful and exciting year in 2011.

Executive Director's report – Ross Wilkinson

At the eResearch Australasia Conference held in November 2010, it was clear that research data infrastructure activity had risen dramatically over the last year. There were a large number of reports of work done at institutions around the country on more effective capture and sharing research data.

It was particularly pleasing to see the cross-fertilisation between projects - we heard about the University of Melbourne working with Griffith University, the University of Southern Queensland working with the University of Newcastle, and there was a real buzz at a very well attended Birds of a Feather meeting organised for participants in ANDS projects to share their ideas. The Australian Research Data Commons is under active construction around the country!



Addressing the pending shortfall in global food production

The UN estimates that, in order to feed the world's population, world food output will need to increase by 70% by 2050. With most arable land already being farmed and existing agriculture threatened by climate change, the challenge is massive. Substantial government and industry investments in recent years have seen Australia make numerous advances in plant genomics and modern breeding technologies, but science has hit a bottleneck in its ability to relate the performance of particular plants to their genetic make-up. Future progress lies in translating the huge database of genomic knowledge into improved agricultural products.

Professor Mark Tester is an ARC Federation Fellow based at the University of Adelaide. He is Professor of Plant Physiology in the University's School of Agriculture, Food and Wine, Director of the Australian Plant Phenomics Facility, and chair of the Research Advisory Group of the Australian Centre for Plant Functional Genomics. His research aim is to elucidate the molecular mechanisms that enable certain plants to thrive in sub-optimal soil conditions, such as high salinity or low water supply. Professor Tester's research is ultimately directed toward addressing the pending shortfall in global food production.

Professor Tester has been instrumental in establishing The Plant Accelerator, a world leading automated plant growth and analysis facility. From the outside, the building looks like a moderately sized glasshouse, built low on the foothills behind Adelaide on the University of Adelaide's Waite campus. It has won several architectural and building industry awards in recognition of its design and water and energy efficiency. This high-tech facility features over 1km of conveyor systems, state of the art imaging, robotic, and computing equipment, all configured to facilitate continuous, automated, and non-destructive measurement of plants' physical attributes (phenotypes). The Plant Accelerator can accommodate and grow 160,000 plants every year, in research programs aimed at increasing drought and salinity tolerance in wheat, barley, grapevines, and other economically important crops.

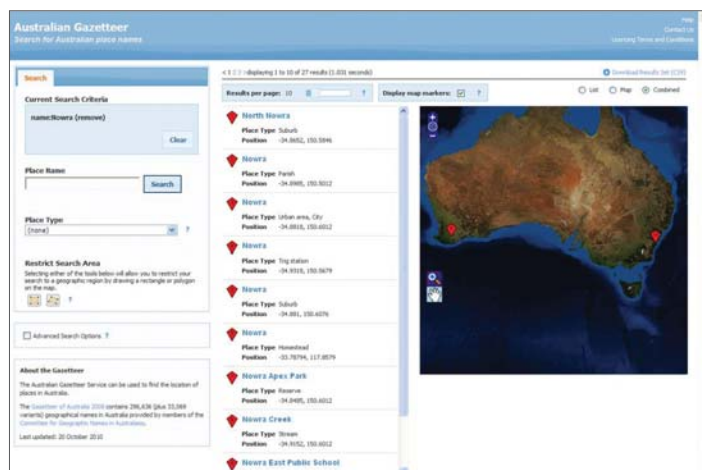
Professor Tester has a vision of an integrated portal to plant genomic, phenomic, proteomic and metabolomic data. Such a portal would provide scientists with a complete bioinformatic picture of all measurable outcomes of plant genomics. ANDS is working with The Plant Accelerator to develop a system that will describe plant phenomics data, then make it visible and discoverable, and ultimately shareable and re-useable through Research Data Australia. The ANDS project at The Plant Accelerator also aims to link in with the Phenomics Ontology Driven Database (PODD). PODD is a National eResearch Infrastructure Taskforce (NeAT) project to develop a web-based phenomics data repository that will capture, manage, secure, distribute and publish raw and analysed phenotypic data. The Plant Accelerator's link with PODD means that Plant Accelerator data will also be discoverable through the specialised PODD phenomics portal. The project is a step on the way towards the phenomics component of Professor Tester's vision.



A Smarthouse at The Plant Accelerator® fitted with automated conveyor systems and imaging stations for the non-destructive phenotyping of plants. Photo courtesy of The Plant Accelerator®, Australian Plant Phenomics Facility.

"Professor Tester has a vision of an integrated portal to plant genomic, phenomic, proteomic and metabolomic data."

MyMaps.gov.au website to be deployed early this year



A screenshot of the prototype version of the MyMaps.gov.au website. Image courtesy of the Office of Spatial Data Management

Researchers and the public need an authoritative source of information about Australian geographic names as this is essential for understanding the social, economic and physical dynamics of communities. Previously, researchers have had to pay to access this data. Ben Searle, General Manager of the Office of Spatial Data Management hopes that a new ANDS funded Gazetteer Service will be fully operational and available early this year. ANDS assisted with the development of the specifications for the web service that will not only be available to the public but will also be machine-readable.

It is to be deployed under a website called MyMaps.gov.au. The digital information is dynamic and it will be possible to download it into a number of different formats. In addition to the place names information it will also have a directory of about 30,000 spatial datasets held by various Australian government organisations using a capability known as the Australian Spatial Data Directory.

The Gazetteer Service differs from other electronic gazetteers because the data is sourced from the legal creators of that data – from the local, state and federal government jurisdictions. In the future the plan is to integrate the Gazetteer Service with other capabilities that will build on the 'MyMaps' environment. The intention is to extend the quantity and richness of the content and to increase the number of datasets that are made available. Ben Searle said, "We want to be able to go beyond simple point data and intend to include various geographic boundaries along the lines of postcodes, electoral boundaries, maritime boundaries or national parks in future versions." Searle added, "Recommendations by the Gov 2.0 Taskforce is one of the drivers that is leading to more open access to spatial data. Freedom of Information reforms have also led to the appointment of an Information Commissioner whose role is to ensure as much government information as possible is made publicly available." MyMaps will assist the government to conform to Gov2.0 and FOI recommendations.

Project Update on Australian Research Data Commons - Party Infrastructure Project

The scholarly information system has always been plagued by the lack of a common public identifier for researchers which can be used to reference them as the authors and contributors of research outputs. This project aspires to address this problem by providing infrastructure to support the implementation of a common public identifier for Australian researchers so that the Australian Research Data Commons can link research datasets to a particular researcher or a research group in spite of changes in name and affiliation over time. This identifier will integrate with global researcher identity systems.

The National Library of Australia (NLA) is expanding its Party Infrastructure (formerly known as People Australia) to allow research institutions to contribute public information about their researchers and research groups (such as names, areas

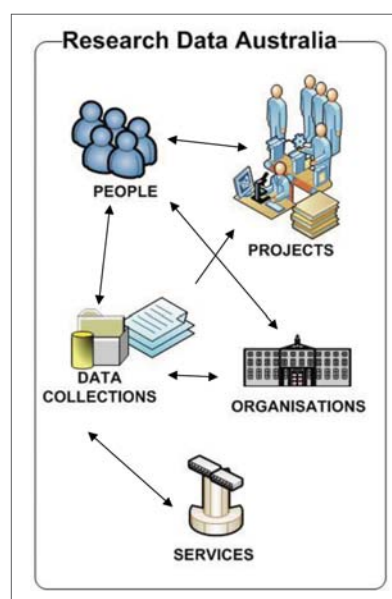
or research, research projects, publications etc) and hence be assigned a unique identifier called the NLA Party Identifier available via Trove (<http://trove.nla.gov.au>) and web services interfaces. Other identifiers such as those assigned to authors by the large academic publishers can be linked to this identifier. Many researchers already have an NLA Party Identifier as they were already included in the Australian Name Authority File which has always been managed by the NLA.

The first phase of this project involved community consultation and the development of a detailed description of how this infrastructure would operate. This was completed in July 2010 and implementation of the infrastructure is now proceeding with completion due in July 2011. A number of early implementers have already provided researcher information to the NLA in order to test the concepts.

The main deliverables of this second and final phase of the project are:

1. better matching rules for researcher identities so that researcher records supplied by institutions to the NLA are more likely to be automatically assigned their NLA Party Identifier
2. an online web-based identity matching interface which will allow them to quickly match researchers against existing entries in the People Australia service and create new identities if required
3. the development of workflow processes along with ANDS to streamline the management of harvesting party information from universities and other research institutions

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



ANDS activities at the eResearch Australasia conference 8-12 November 2010



ANDS staff at eResearch Australasia

ANDS had a strong presence at the 2010 eResearch Australasia conference, held on the Gold Coast between 8 and 12 November. The ANDS presence was most obvious on the Wednesday, when all ANDS staff wore ANDS t-shirts, along with other conference delegates working on ANDS-supported projects at their institutions. Many conference presenters commented on the size of the ANDS tribe. ANDS also had a high profile in the content of many of the conference presentations, and acknowledgements of ANDS funding were frequently heard during presentations on a wide range of eResearch topics.

Issues around research data were addressed in many conference presentations. In every parallel session at least one, often two, and sometimes three of the four streams featured presentations around



eResearch Australasia plenary session on 8 November 2010

various aspects of research data. Observations of the need for good metadata and of the growing scale of the problem of research data management were common.

The ANDS Birds of a Feather session was attended by about 100 people who are involved in ANDS funded projects. It was a great opportunity for members of the wider ANDS community to come together and share information about their achievements. Sam Searle (Monash), Robyn Rebollo (Griffith), Craig Milne (QUT) and Gerry Ryder (CSIRO) gave short presentations on their work to date and took questions from the floor. A key characteristic of this session was the willingness shown by the ANDS partners to make connections and to recognise opportunities for working together in the future.

Data Sustainability: applying archival practice to research data Workshop

Data Sustainability: applying archival practice to research data was the title of a workshop sponsored by ANDS and held in association with eResearch Australasia 2010. The workshop was led jointly by Dr Andrew Wilson, previously of ANDS but now working with the Queensland State Archives, Gavan McCarthy from the University of Melbourne, and Stuart Hungerford, now working with ANDS but previously with the Supercomputer Facility at the Australian National University. About twenty people attended, from both university and public sector backgrounds.

The workshop provided an overview of archival principles and practices for sustainability and related these to the requirements of those needing to curate research data. Having looked at the principles, the session moved on to PARADISEC as a case study, prompting an energetic discussion which concluded the afternoon.

ANDS contributed to the poster sessions at eResearch Australasia with its own poster on building a culture of data citation. This poster was designed to demonstrate the work being done by ANDS to develop an infrastructure which will support the growing need for data to be shared and rewarding researchers for doing so.

A copy of the poster is available at

http://ands.org.au/guides/data_citation_poster.pdf

Forthcoming events

Queensland ANDS Outreach community event:

17th February 2011, Brisbane at QUT

Contact: Andrew White (andrew.white@ands.org.au)

South Australia ANDS Outreach community event:

24th February 2011, Adelaide at Art Gallery

Contact: Andrew Williams (andrew.williams@ands.org.au)

Stop Press

Congratulations to Sam Searle, Data Management Coordinator at Monash University, who received the 2010 Council of Australian University Librarians (CAUL) Achievement Award for leadership in data management.

For more information see: <http://ands.org.au/news/caulachievementaward2010.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 Office
C/- Monash University
Clayton, VIC 3800
Telephone: 03 9902 0585
Email: contact@ands.org.au

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

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

ANDS Project Partners:



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/>