AUSTRALIAN NATIONAL DATA SERVICE (ANDS)
Australian Research Data Commons
Education Investment Fund (EIF)

Annual Report 4, 30 September 2013
# TABLE OF CONTENTS

1 PROJECT STATUS ................................................................. 4  
   1.1 Background ........................................................................ 4  
   1.2 Major Activities, Breakthroughs, Highlights, and Issues .................... 7  

2 ACTIVITIES UNDERTAKEN .......................................................... 11  
   2.1 Research Infrastructure ......................................................... 11  
   2.2 Data Capture ........................................................................ 12  
   2.3 Metadata Stores ................................................................... 18  
   2.4 Public Sector Data .................................................................. 24  
   2.5 ARDC Core Infrastructure ....................................................... 30  
   2.6 ARDC Applications ................................................................. 46  
   2.7 International Infrastructure ..................................................... 49  
   2.8 Project Office ........................................................................ 51  
   2.9 Promotion ............................................................................ 52  
   2.10 Risk Management .................................................................. 58  

3 PROGRESS AGAINST MILESTONES ............................................. 59  
   3.1 Data Capture Infrastructure ..................................................... 59  
   3.2 Metadata Store Infrastructure .................................................. 60  
   3.3 Public Sector Data Infrastructure .............................................. 61  
   3.4 ARDC Core Infrastructure ....................................................... 62  
   3.5 ARDC Application Infrastructure .............................................. 63
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6 International Infrastructure</td>
<td>63</td>
</tr>
<tr>
<td>4 DEVIATIONS FROM THE PROJECT PLAN</td>
<td>65</td>
</tr>
<tr>
<td>5 COMMITMENTS</td>
<td>66</td>
</tr>
<tr>
<td>6 FINANCIAL AND HUMAN RESOURCES</td>
<td>67</td>
</tr>
<tr>
<td>7 CO-INVESTMENT</td>
<td>69</td>
</tr>
<tr>
<td>7.1 Access and Pricing</td>
<td>69</td>
</tr>
<tr>
<td>7.2 Project Co-Investment</td>
<td>69</td>
</tr>
<tr>
<td>8 PERFORMANCE INDICATORS</td>
<td>70</td>
</tr>
<tr>
<td>8.1 KPI Report</td>
<td>70</td>
</tr>
<tr>
<td>8.2 Progress over the Life of the ANDS Project</td>
<td>73</td>
</tr>
<tr>
<td>8.3 Overall Progress</td>
<td>74</td>
</tr>
<tr>
<td>9 AUDIT STATEMENT</td>
<td>76</td>
</tr>
<tr>
<td>10 APPENDICES</td>
<td>77</td>
</tr>
<tr>
<td>10.1 Confidential Information</td>
<td>77</td>
</tr>
<tr>
<td>10.2 Project Description Detail</td>
<td>77</td>
</tr>
<tr>
<td>10.3 Progress against activities</td>
<td>89</td>
</tr>
<tr>
<td>10.4 Risk Register</td>
<td>109</td>
</tr>
</tbody>
</table>
1 Project Status

1.1 Background

Research data is simultaneously becoming more voluminous, more complex, and more vital as the very nature of research changes. Research has become more investigative because it is possible to assemble significant data collections that enable much broader problems to be addressed. Thus it is essential that research data is managed, able to be assembled, connected to other data and used to address problems that may well be different to the reasons for gathering the data in the first place. The Australian Government recognised the significance of data as key infrastructure in supporting research excellence and research innovation when it established the Australian National Data Service (ANDS).

ANDS has been in operation since January 2009 as part of the NCRIS initiative. Its aim of enabling more researchers to reuse research data more often required establishing partnerships beyond ANDS, and this need is continuing to increase. In May 2010, the Australian Research Data Commons (ARDC) project was announced as an EIF funded Super Science Initiative, to be managed by ANDS. An agreed Project Plan was submitted in June 2009 and accepted in September 2009. Some activity in the NCRIS funded ANDS project was transferred to the ARDC project as a result. It was subsequently agreed that ANDS operations should be extended beyond June 2011, to June 2013. In October 2012, the Collaborative Research Infrastructure Scheme (CRIS) was announced, and ANDS received an additional $3M to maintain minimal critical infrastructure. This led to a substantial modification of the submitted 2012-13 Annual Business Plan. Each of these changes has had a high impact on the activities of ANDS, and the second change has substantially affected the ANDS project as agreed in the 2012-13 Business Plan. As each of these changes has occurred, ANDS has continued to manage the ANDS and ARDC projects together, as they are strongly co-dependent. This report describes activity taking place against the agreed 2012-13 Business Plan, subject to the agreed variation to the plan submitted in March that described the effect of the business plan variation.

At the time of the submission of the 2009-10 Business Plan, ANDS had four programs of activity:

- **Developing Frameworks** – the frameworks that will enable research data producing institutions to capture, manage and share research data.
- **Providing Utilities** – services that reduce the cost of capture and ease the task of discovery.
- **Seeding the Commons** – improving local data capture and populating the data commons.
- **Building Capabilities** – improving Australia’s capability to manage its research data.

As a result of the ARDC project, the NCRIS ANDS project was consolidated into two programs of activity:

- **Frameworks and Capabilities** – the frameworks that will enable research data producing institutions to capture, manage and share research data; and improving Australia’s capability to manage its research data.
- **Seeding the Commons** – improving local data capture and populating the data commons.
The associated ARDC project has five programs of activity:

- **Data Capture** – an institutionally based program to automate the capture of data and metadata from instruments (broadly defined) in data intensive research.
- **Public Sector Data** – a program of making more public data collections visible and available through the ARDC.
- **Metadata Stores** – an institutionally based program that enables metadata to be stored coherently across an institution that supports data management, publishing, sharing and reuse.
- **ARDC Core Infrastructure** – an ANDS driven program that puts in place the national services that enable research data to be published and discovered (it is an expansion of the *Providing Utilities* program).
- **ARDC Applications** – a program that develops tools and services to support demonstrations of the value of exploiting data in the ARDC.

Two new programs were subsequently created in 2012:

- **National Collections** – an ANDS-driven, NCRIS-funded program partnering with institutions wishing to make National Collections available, and with RDSI and its nodes to help improve storage and access to those collection.
- **International Infrastructure** – a program designed to work collaboratively with international organisations and partners to ensure a more compatible international data-sharing environment for Australian researchers.

These programs were created in response to the changing environment, rather than the changing focus of ANDS. The advent of the RDSI initiative meant that there was new opportunity to assemble and make available collections of significance to researchers, to research disciplines, to research institutions and to the nation. ANDS’ relationship with institutions meant that ANDS could focus effort to ensure that more collections of strategic significance were available on RDSI nodes where the collections are managed, connected, discoverable and increasingly usable, through better access, and possibly with new forms of access.

The strong drive in Europe and the US in particular, but also more generally, to create an environment that enables data to be shared across boundaries, provided Australia with an excellent opportunity to engage internationally, particularly through the newly forming Research Data Alliance. This international initiative should enable Australian researchers to partner more effectively through a shared research data environment.

Figure 1 shows how the NCRIS programs complement and inter-relate to the creation of the Australian Research Data Commons.

Taken together, the intent of the two investments is to:

- Create an “essential meeting place where the Australian path forward for research data management can evolve and where a vision can be achieved.” (*Towards the Australian Data Commons*), developed during 2007 by the ANDS Technical Working Group.
- Enable the following capability: “Research data and research outputs from all sources can be discovered and reused across disciplines and over time through an integration of repositories and data centres supporting national and specialist discovery services.” (Towards the Australian Data Commons).

- Create and populate the Australian Research Data Commons, which “will support the discovery of, and access to, research data held in Australian universities, publicly funded research agencies and government organisations for the use of research.” (ARDC Investment Plan).

![Diagram: Relationship between Programs]

**Figure 1: Relationship between Programs**

The ARDC investment plan quoted above can thus be seen as an intensification of effort in support of the second Towards the Australian Data Commons statement.

The result of the CRIS investment was that ANDS has undergone a further change and simplification of its program structure. Its current programs are National Services, National Collections, Institutional Engagements and International Collaboration. This structure was used again to determine a response to the NCRIS 2013 Scheme that is beyond the scope of this report. However, whilst this simplification was gradually bought into effect late in 2012-13, it was felt that as the bulk of the activity of ANDS was conducted under the structure depicted in Figure 1, and to ensure that the activities of ANDS can be described against the agreed Annual Business Plan, this report should use the structure depicted in Figure 1.
1.2 Major Activities, Breakthroughs, Highlights, and Issues

Overarching achievements, some of which have been reported in previous Annual Reports, include:

 The Australian Research Data Commons (ARDC) has been established, and substantial progress has been made in populating it. The ARDC is a combination of the set of shareable Australian research collections, the descriptions of those collections, including the information required to support their reuse, the relationships between the various elements involved (the data, the researchers who produced it, the instruments that collected it and the institutions where they work), and the infrastructure needed to enable, populate and support the commons. In summary, all components of the ARDC exist, but not all components have been established and used at all relevant institutions, or by all relevant researchers.

 ANDS has driven a change in the research data management uptake in Australia. ANDS is engaged with all major research institutions, and importantly they are engaged with and learning from each other’s approaches.

 Research data infrastructure and research data management have been established at a significant number of research institutions. ANDS estimates that there are approximately 300 people working on data management within research institutions, which is probably a ten-fold increase compared to January 2009. Research institutions are seeing substantial value in this infrastructure (see for example quotes from the University of Adelaide and James Cook University). ANDS investments at institutions have triggered substantial co-investment and post-investment, with over $2M of institutional investment made to date, and over $3M of post-project investment. This indicates the extent to which institutions are embedding a research data infrastructure into standard operations.

 Data is overwhelmingly on the agenda in research and research infrastructure, and ANDS has helped position Australia internationally. The Research Infrastructure Roadmap produced in 2011 saw data as crucial infrastructure for research, and this has been emphasised in the draft National Research Investment Plan, which refers to the crucial role information and data play in enabling Australian research to tackle the key research challenges of the country. Very importantly, research institutional leaders are similarly seeing great value in the way that research data can give their researchers an advantage in research data partnerships and tackling larger research questions. This perception is mirrored internationally. There was a very strong emphasis on research data in the 2012 International Conference on Research Infrastructure in Copenhagen. Both Europe and the US have made significant investments in research infrastructure. ANDS has had an important role in ensuring that Australia has a leading role in international research infrastructure initiatives, particularly through the emerging Research Data Alliance, where Australia is partnering with
the US and the EU. ANDS, together with some of the data-intensive capabilities, has made a significant contribution in ensuring that Australian researchers and research institutions are engaged and leading in these global trends.

- The Australian Research Data Commons has matured and grown substantially. There are twice as many collections and three times as many contributing institutions as compared to last year, and the ARDC now covers every Field of Research. Importantly, now that the Data Citation service is available, researchers are availing themselves of the opportunity to publish their research data, using minted Digital Object Identifiers to connect their collection description to the relevant data repository, and to cite their data.

The 2012-13 business year was extremely busy for ANDS, with many activities underway. It saw no major change in direction or approach, however there was considerable change in approach as it became clear that ANDS needed to slow expenditure to ensure there was no reduction in ANDS services below “minimal critical” levels. While managing the substantial set of activities, ANDS confronted a number of issues that it had to deal with and learn from:

- There was a continuing challenge for ANDS working with a large number of partners delivering a large number of data infrastructure components simultaneously around the country.
- ANDS has also needed to transition beyond project support to institutional partnerships, where there was significant focus on other parts of the eResearch landscape through the increasing activities of RDSI and NeCTAR in particular. The ANDS approach of focusing on institutional partnerships has contrasted with other approaches, but has remained a focus given the critical role of research institutions in research data infrastructure.

The major activities for ANDS as a whole over the 2012-13 period were:

- Making more than 56,599 research data collection pages discoverable through Research Data Australia, Google, and other search engines that harvest ANDS, from over 70 institutions providing research data.
- Substantially updating the national data services, enabling better publication and discovery of research data.
- Over 30,000 archival collections from major state archives are now discoverable through Research Data Australia.
- All major research institutions have installed substantial research data infrastructure, notably metadata stores, that enable them to participate in the Australian Research Data Commons.
- Demonstrating the value of reusing data with a suite of applications that has enabled high profile researchers to provide compelling demonstrations of new ways of conducting data intensive research.
- Establishing the Research Data Alliance -- in partnership with the EU and the US -- which is committed to “data sharing without barriers”, holding an initial Plenary in Gothenburg, Sweden, and attracting a very substantial level of international commitment.

Taken together, two important changes have occurred during the year. Firstly, there is clearly an increased desire to publish and get credit for research data. The great interest in data citation is evidence of this, so...
the ability of Australian researchers to publish their data will be increasingly important. Secondly, Australia’s response to the open data agenda internationally has been made easier by having a locus of thinking about research data and having conversations, developing policy responses, and involving all relevant agencies, through ANDS policy and practice capabilities.

There have again been many highlights over the past year; perhaps they can be best reflected by describing six examples of research impact.

**Edgar - James Cook University**

The Edgar website allows visitors to explore the potential impact of future climate change on Australian birds. Edgar shows locations where a bird species has been observed, and displays current and future species ranges for Australian birds under multiple climate change scenarios. Edgar enables scientists, the public, and policy-makers to see what is likely to happen to an Australian bird species in the future and can therefore assist them in making decisions regarding conservation and climate change action.

Edgar uses bird observation records from the Atlas of Living Australia database to generate the current and future species distribution models. Obviously, errors creep into the data and so the website allows other experts and birdwatchers to help improve the accuracy of Edgar’s projections by classifying observations, triggering re-running of models.

This is a giant leap in communication of both the potential impacts of climate change on Australian birds and the value of models for projecting likely impacts of climate change. For more information see [http://tropicaldatahub.org/goto/Edgar](http://tropicaldatahub.org/goto/Edgar).

**TissueStack – University of Queensland**

The Centre of Advanced Imaging has developed a web-based 3D viewing system, called TissueStack, for large imaging datasets. The novelty of TissueStack is its ability to link to specific parts of the data, and rapidly view and collaboratively annotate on very large 3D datasets via a web browser. As a result, the project has filled the current research data sharing gap between those in imaging centres and laboratories that acquire high-resolution data, and those who use these data in their own research.

TissueStack has greatly enhanced research collaboration by enabling researchers around the world to access and annotate the same dataset from a cloud and from a convenient device – desktop computer, tablet or mobile phone. TissueStack was originally designed for use in the medical and scientific 3D imaging field. However, the viewing system has since been used for multiple datasets including museum samples, coal mining samples and electron microscopy data. Perkin Elmer are now proposing to provide this software with the instruments they manufacture. For more information see [http://q0013-webdav.qcloud.qcif.edu.au/desktop](http://q0013-webdav.qcloud.qcif.edu.au/desktop).

**Soils to Satellites (S2S) – Atlas of Living Australia (ALA) and TERN (Terrestrial Ecosystem Research Network) Eco-informatics**

The S2S web tool was developed to bring together biodiversity, genomic and environmental data from a number of different source organisations, including ALA, TERN’s AusPlots, TRENDSA (Transect for Environmental Monitoring and Decision Making), AusCover facilities and Bioplatforms Australia.

“Associate Professor Jeremy VanDerWal, James Cook University, says the help of ANDS was “absolutely vital” to the development of ‘Edgar’, a web portal exploring the potential impact of future climate change on Australian birds. “I started this project developing it myself – I’m not a developer, I am a researcher. ANDS has enabled me to take what I had as an idea and bring it through to something that is functional that people can actually use. Without [ANDS] it would have never gotten to that stage.”
S2S provides a single interface that enables researchers and other users to visualise relationships and patterns across these different data sources, and to create and download datasets for further reuse. S2S creates research efficiencies, which means more effort can be devoted to advancing an understanding of how ecosystems work and help inform environmental decision-making. For more information see http://soils2sat.ala.org.au/.

Indigenous Australian Rock Art Data System – University of Western Australia

The Centre for Rock Art Research and Management at the University of Western Australia has developed its Rock Art Data System as a repository for its rock art research data collections. It provides contextual information about research grants, associated researchers and other key metadata. It also seeks to protect Indigenous rock art by advocating, through these data collections, its heritage value.

Its key objectives are to improve the maintainability of archaeology datasets at the University of Western Australia, to make the data available to Indigenous Australian communities, and to share data with collaborating researchers and also with the wider research community. This enables similar indigenous research projects around the world to examine various contexts and aspects associated with rock art, from both indigenous and non-indigenous perspectives. This supports the belief that research can achieve a balance in the efforts to use indigenous people’s visual heritage to learn more about its role in the lives of indigenous communities. For more information see http://researchdataonline.research.uwa.edu.au.

Founders and Survivors: Genealogical Connections – the Australian National University and of the University of Melbourne

The Genealogical Connections project – part of the University of Melbourne’s large-scale Founders and Survivors project – brings together research datasets created from the records relating to the 73,000 convicts transported to Tasmania in the nineteenth century and their descendants. The datasets include convict lists, birth, death and marriage registrations, pauper and hospital admissions and World War One service records.

The project has developed a population database of national and international significance for historical, demographic and population health researchers. The project has developed software allowing fully flexible linkages of individuals to larger family groups, and extraction of these data for analysis and visualisation. For more information see http://foundersandsurvivors.org/.

SMART Infrastructure Dashboard – University of Wollongong

In many developed countries, including Australia, infrastructure services (electricity, water, gas, sewerage and waste collection) are increasingly provided by a complex set of public and private agencies. The SMART Infrastructure Dashboard (SID), a customer-centric web portal, uses Geo-Business Intelligence to provide an integrated view of these services for better governance and productivity. SID not only provides a much-needed planning and policy support tool – to help cities and regions with long-term, strategically planned infrastructure – but also pushes the boundaries of Geo-Business Intelligence beyond traditional use. For more information see http://smart.uow.edu.au/.
2 Activities Undertaken

2.1 Research Infrastructure

ANDS has continued to make progress towards its goals of providing greater support to enable researchers to work in the new world of data-intensive research, notably through the substantial increase of effort in describing data collections and making the descriptions automatically visible through Research Data Australia (RDA), the increased use of specific services (Identify My Data and Register My Data), roadshows and “boot camps” to improve the capability of institutions to manage and share their research data, engagement with specific institutions to better support their data management, and the provision of advice and documentation in various data management areas via the ANDS website. More detailed reports on progress in this area are contained in section 10.2. The effort on infrastructure development can be seen in Figure 2 where the dark green pipes and green boxes show the infrastructure being created in the ARDC project.

Complementary infrastructure established as part of the ARDC project is described in a separate progress report. Some of this infrastructure was to be established in the NCRIS ANDS project, but as a result of the ARDC project plan and a modified ANDS Business Plan, the work was undertaken under the ARDC project.

The overall infrastructure is being constructed through 5 programs:

- **Data Capture** is constructing the pipes that connect data sources to the data stores and the metadata stores
- **Public Sector Data** is connecting data held in public sector agencies to the commons either from their data and metadata stores to the ANDS portal
- The **Metadata Stores** program is creating a set of metadata store solutions that can be deployed at research institutions
- The **ARDC Core** program is creating the infrastructure that enables collections to be identified, harvested and discovered through the ANDS portal
- The **ARDC Applications** program is designed to enable researchers to exploit the whole of the ARDC infrastructure to get new value from existing data.

A new program focused on **International Infrastructure** was introduced during the year to consolidate international activities and reflect an increasing focus on International Infrastructure. This program was funded by a combination of EIF funds and funding from separate contracts with the Department of Infrastructure.

The next section describes the specific research infrastructure created in the 2012-13 financial year.
2.2 Data Capture

2.2.1 Overview of program

The Data Capture program aims to simplify the process of researchers routinely capturing data and rich metadata as close as possible to the point of creation, and depositing these data and metadata into well-managed stores. Metadata will need to be held at both collection and object level in order to support reuse.

The Data Capture program will achieve this aim by augmenting and adapting existing data creation and capture infrastructure commonly used by Australian researchers and research institutions to ensure that the data creation and data capture phases of research are fully integrated so as to enable effective ingestion into the Research Data and Metadata Stores at the institution or elsewhere. This integration will make it easier for researchers to contribute data to the ARDC directly from the lab, instrument, fieldwork site, etc. It will also ensure that higher quality metadata (critical for reuse and discovery) is produced through automated and semi-automated systems. The approach taken will be to partner with leading research groups and Super Science initiatives to augment or adapt data creation and capture systems.

The resulting infrastructure components will include software to integrate tightly with the experimental environment of the researcher to take the data that is being captured/created, and augment this with metadata that describes the setting within which the data is being captured/created, as well as other relevant details (where available) about the research project, researcher, experiment, sample, analysis and instrument calibration details. ANDS will also adopt/adapt/develop software to facilitate automatic/semi-automatic deposit from instruments into data stores/repositories.

The Data Capture program was originally allocated $12M in the EIF ARDC Draft Project Plan. Following the process of public consultation around this Draft Project Plan, this amount was increased to $18.47M. The consultation process also validated the decision to take an institutional approach in allocating the bulk of the available funds. An analysis of research intensity for the major Australian research-producing institutions was undertaken in late 2009 based on the most recent publicly available data on research productivity, and $11.6M of Data Capture funds was allocated in bands of $1M, $500K, or $200K. In late 2009 institutions were each sent an individual invitation to take part in an Expression of Interest process.

2.2.2 Outline of projects

An analysis of research intensity for the major Australian research-producing institutions was undertaken in late 2009 based on the most recent publicly available data on research productivity, and $11.6M of Data Capture funds were allocated in bands of $1M, $500K, or $200K. In late 2009 institutions were each sent an individual invitation to take part in an Expression of Interest (EOI) process.

At June 30, 2012, ANDS had contracted and commenced all of the Data Capture projects at all of the EOI institutions. A breakdown of the progress made for these projects over the reporting period is provided below.
<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Title</th>
<th>Project agreed</th>
<th>Contracted</th>
<th>Under way</th>
<th>Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Australian National University</strong></td>
<td>Earth Sciences</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Optical Astronomy (Skymapper)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Phenomics</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Humanities and allied disciplines</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>ANSTO</strong></td>
<td>Metadata management for neutron beam instrument data - a joint project with</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Australian Synchrotron</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metadata Capture and Storage for the Three Mature Beamlines at the</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Australian Synchrotron</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CSIRO</strong></td>
<td>Research Data Service: Multi Source (incl. Sensor Network) Data Capture</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>ANDS-CSIRO-ATNF Pulsar Data Management Project</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Curtin University of Technology</strong></td>
<td>Curtin deployment and configuration of Institutional Metadata Repository and</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Research Data Portal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deakin University</strong></td>
<td>Filtration Membrane Fouling Data Collection for Water Treatment Research</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Crystal Orientation Data Collection for Conversion to a General Data Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Flinders University</strong></td>
<td>Automated measurement of the responses of wildlife populations to climate</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Griffith University</strong></td>
<td>Smart Water</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Adult Stem Cell &amp; Neurobiological Microscopy Instrumentation and Research</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Data Management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>James Cook University</strong></td>
<td>Tropical Data Hub</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>La Trobe University</strong></td>
<td>CMSS RLI Metadata Capture and Publication</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Macquarie University</strong></td>
<td>Glycomics Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Papyri Data Capture</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Project Title</td>
<td>Project agreed</td>
<td>Contracted</td>
<td>Underway</td>
<td>Completed</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Monash University</td>
<td>Research Data Management of the Monash Weather &amp; Climate Program (Climate and Weather)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Biomedical Data Platform (Molecular Biology)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Tools for curating and publishing research data in the form of media collections (Multimedia Collections &amp; ARROW)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Capture and publication of Australian ecosystem data from a network of measurement sites (Ecosystem Measurements)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Capture and publication of data on the history of adoption (History of Adoption)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Data Publication to Interferome (MIMR/Interferome)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Comprehensive Data Management for Microscopy Research Datasets</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Queensland University of Technology</td>
<td>Greenhouse Gas Emissions from Australian Soils</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Biodiversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B150 Big Jam</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RMIT University</td>
<td>Data Capture from High Performance Computing Multi-User Environments</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Adelaide</td>
<td>Genomics Data Capture</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Automated capture and publishing of data generated on high throughput plant phenomic platforms.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Melbourne</td>
<td>Melbourne Neuropsychiatry Centre (MNC) Bioinformatics Development Project</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Youth Research Centre’s Life Patterns Project: Longitudinal qualitative and quantitative survey data capture and reuse</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Video data in the Social Sciences. Optimising Metadata Capture, Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Title</td>
<td>Project agreed</td>
<td>Contracted</td>
<td>Under way</td>
<td>Completed</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Sharing Procedures and Long-term Reuse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Federated Neuroimaging Collections in the National Data Commons</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Humanities and Social Science Research Data at the University of Melbourne</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Capture of Complex Data to Support Clinical Research in Cardiovascular and Neurological Medicine</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Founders and Survivors Project</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Enhanced Metadata Capture for Sustainable Management, Sharing and Reuse of APN Histopathology Research Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of New South Wales</td>
<td>ARDC Linked International Glycomics Repository &amp; Instrument Data Capture</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>An international antibiotic-resistance gene cassette database</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>ANZNN Neonatal Data Capture Portal</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data capture and integration across multiple platforms</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Managing and Sharing Genomic Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>Data Capture for the Data Commons</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Queensland</td>
<td>Spatially Integrated Social Science</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Microscopy/Microanalysis Image and Data Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>DIMER Diffraction Image Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Aquatic Species Tracking Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>3D Anthropological and Archeological Collection Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>The Health-e-Reef Project</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Linking the EMBL Australia EBI Mirror with the Australian Research Data Commons</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>Development and testing of a data capture</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Title</td>
<td>Project agreed</td>
<td>Contracted</td>
<td>Under way</td>
<td>Completed</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>South Australia</td>
<td>tool for instruments at the Ian Wark Research Institute</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Sydney</td>
<td>SKAMP Data Capture: astronomy (NOTE: this has morphed into SHED – Sydney Harbour Estuary Data)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NSW TARDIS Node</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AgDataCapt: Capturing Agricultural Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Metadata Store/Aggregator</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExCite9: a workflow and tools for improving fieldwork data collection and submission to institutional repositories</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Clarke eHealth (Early Activity): Capture, management, reuse and discovery of breast cancer microscopy virtual images</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Redmap Australia</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>University of Tasmania</td>
<td>Data Capture of state-wide hydrological datasets</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Maximising the Benefit from Data-Intensive Processes at UTS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>University of Technology, Sydney</td>
<td>Deployment and configuration of Institutional Metadata Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>University of Western Australia</td>
<td>Integrated Data Capture for Characterization and Analysis</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Archaeological Rock Art Data Capture</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marine Ecology Video Capture and Storage</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Climate Change and Energy Research Facilities (CCERF)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>University of Western Sydney</td>
<td>Biomechanics Data Capture Project System</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>University of Wollongong</td>
<td>Remote Sensing Spectral Library</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Status of Data Capture projects
2.2.3 Activity/Deliverables for 2012-13

The following projects were completed during the reporting period. Software and other deliverables are made available from the ANDS projects registry: http://projects.ands.org.au/.

- ANU – Earth Sciences
- ANU – Optical Astronomy
- ANU – Phenomics
- ANU – Humanities and allied disciplines
- CSIRO – Research Data Service Multi-Source Data Capture
- Curtin – Institutional Metadata Repository and Research Data Portal
- Deakin - Crystal Orientation Data Collection for Conversion to a General Data Type
- Deakin - Enhancing Filtration Membrane Fouling Data Collection for Water Treatment Research
- James Cook University – Tropical Data Hub
- Macquarie – Glycomics Repository
- Monash – Comprehensive Data Management for Microscopy Research Datasets
- Adelaide – Genomics Data Capture
- UNSW – Data Capture and Integration
- Newcastle – Data Capture for the Data Commons
- University of Queensland – Spatially Integrated Social Science
- University of Queensland – Microscopy/Microanalysis Image and Data Repository
- University of Queensland – DIMER Diffraction Image Repository
- University of Queensland - 3D Anthropological and Archeological Data capture of 3D digital models
- University of Sydney – Metadata Store Aggregator
- University of Sydney – ExCite9
- University of Tasmania - Redmap
- UTS – Maximising the benefits from data-intensive processes at UTS

For a more detailed description of the projects please refer to Appendix 10.2.1

2.2.4 Program Highlights, Issues and Breakthroughs

The main highlight of the previous year has been the completion of a large number of projects. This has resulted in the delivery of a wide variety of software from across most research areas, and in the description and dissemination of many hundreds of research data collections. These projects have delivered infrastructure that supports the ANDS transformation goals, as well as offering substantial material to the product catalogue.

The Data Capture program has had to deal with many of the same challenges that have been experienced in other ANDS Programs: starting the engagement, agreeing on proposals, and finalising contracts. Getting the right level of engagement with institutions in order to start having the conversations about what is possible has often been difficult. Once the conversation has started, getting to agreed proposals has also often required many iterations. Finally, there have been delays in getting to finalised contracts once agreement has been reached on what each proposal involves. ANDS has identified these issues and put in place
improved processes to reduce their occurrence and impact. Partners have also found that finding suitable staff has also presented challenges.

Assessing and monitoring the number of projects in a way that is both auditable and yet does not present a roadblock for the partners is an issue, although considerable work has been done on the processes to refine this.

In addition, work on these projects has often been slower than expected, for a number of reasons, including difficulty in obtaining sufficient researcher engagement, staff recruitment and turnover, partners not understanding ANDS’ requirements and the complications that come from software development. ANDS has worked closely with partners to try and keep projects on track, and we are confident that these issues can be overcome, and the projects will be finished by the ANDS completion date.

More recently the NeCTAR funding rounds have distracted many of our partners from their ANDS projects. In a number of institutions the same staff who have been working on ANDS projects were asked to participate in the NeCTAR applications, which was given a higher priority. This has delayed some ANDS projects as a result. ANDS has also caused a similar problem for itself with the Metadata Stores projects, as these were also handed to the same people doing the work on existing projects. The 2012 Excellence in Research Australia (ERA) exercise is expected to have a similar impact at some institutions. It is clear that resources to work on eResearch in general remain constrained across the sector, with a consequent churn of available staff.

2.2.5 Program Learnings

ANDS projects are often complicated, and require a range of skills to complete effectively. Many institutions underestimated this, and did not resource themselves efficiently. ANDS did not always watch this closely enough, and we were often content to accept assurances that all was well.

Learning: Projects need more advice on what they will need to do up front, and more guidance on what they sign up for to ensure they don’t over commit. Additionally, they need to be watched carefully to ensure they are making steady progress, without having them feel like they are being over-regulated. Finding the right balance can be tricky.

2.3 Metadata Stores

2.3.1 Overview of program

Information that can be held about data (often called metadata) can be grouped into four categories. The first is information for discovery, and is primarily held at the level of a collection. This consists of the range of pieces of information that will assist in the discovery of the collection. The second is information for determination of value (also primarily at collection level). This includes information such as the name of the researcher, institution or funding program that might help a potential user to decide whether they want to access the data. The third is information for access that might be a direct link to the data objects (stored elsewhere, such as on national and institutional data stores), both at collection and possibly object level, or
contact details for where to source the data. The fourth kind of information is *information for reuse*, and will include things like reading scales, field names, variables, calibration settings that are needed in order to effectively reuse the data. This will mostly be at object level.

In practice, the distinction between data and metadata can be somewhat arbitrary and depends on the system that is being used to manage the data. If this system is files-oriented, then the metadata will almost always be separately managed in some sort of associated system. If data management system is database-oriented, then much of the metadata will either be attributes of rows and columns for the database tables.

ANDS is concerned with information about data collections and data objects, but importantly also with information about associated entities. These include parties (both people and organisations), activities (that produce the data) and services (associated with the capture of, and access to, the data). These associated entities serve as part of the rich context for the data collections, and also contribute to the information for discovery and information for determination of value. This rich context is coming from existing institutional systems via software infrastructure that might be thought of as pipes along which the contextual information flows. There also pipes between metadata stores and data stores, and between metadata stores and the ARDC Core infrastructure.

So, software that is being developed or deployed by the Metadata Stores program needs to support a range of functions for different kinds of objects. The first is the creation and management of these kinds of information, or their harvesting from other sources (research management systems, human resources systems, finance systems). In addition, the software needs to manage the relationships between the information about data collections/objects and the data collections/objects themselves. The software may need to support queries over the data by users within the institution. Finally, the software needs to be harvestable by ANDS services, as well as by other organisations. This program therefore needed to help research producing institutions develop, configure and make available this metadata infrastructure.

The required functions can be provided in a wide variety of ways, and via different configurations of software components. In practice, a small number of design patterns are appearing, in part because of the ways in which ANDS has been funding activity at institutions, involving four kinds of stores:

- **Combined Stores**: manage both Collections and Object Metadata for a single institution across a range of disciplines.
- **Collection Stores**: manage the information about data collections within an institution; may also accept feeds from enterprise systems (some of which ANDS has funded), and also feed the ANDS Data Collections Registry.
- **Instrument Stores**: tightly coupled to particular instruments or clusters of instruments. A significant number of these have been developed, not with Metadata Stores funding, but with Data Capture funding. These solutions either feed the ANDS Collections Registry directly (the commonest pattern), or via an institutional Collections store (much less common).
- **For some disciplines**, there are well-established international practices for managing data and metadata, as well as associated software. These Discipline Store solutions might be deployed within institutions or at national or international data centres. ANDS might fund some pipes from instances of these to institutional Collection stores.
As well as these different kinds of metadata stores, the data itself needs to be stored somewhere. This might be a local store (either just associated with an instrument or institutionally supported), one of the offerings that might be made available through RDSI, or an international disciplined-focusressed data store (such as the PDB or EMBL/EBI).

### 2.3.2 Outline of projects

Institutions have many different systems to manage research. To accommodate that variation, ANDS has funded three Combined Stores solutions from the Metadata Stores program. The CSIRO Data Management System has completed its initial development and CSIRO is now extending and progressively generalising it to the entire range of research areas within CSIRO. The Squirrel system has been built by Monash University to meet some of its institutional metadata store needs, managing both object and collection metadata. The development of this system was co-ordinated with the combined metadata store activities at the Australian Synchrotron and ANSTO, and was built on the same codebase.

ANDS has also funded two Collection Stores from the Metadata Stores program. The VITRO solution was developed by Melbourne University (based on the VITRO software and ontology from Cornell) and was deployed by QUT and Griffith University as the basis for their Research Metadata Store Hub. The ReDBox solution was developed by QCIF and was initially deployed by the University of Newcastle.

ANDS is funding nearly eighty Data Capture projects. Many of these are in institutions that at project start had no suitable local metadata solutions and so approximately 35 are creating their own Instrument Stores to capture object metadata and provide collections descriptions to ANDS (and of these 35, 25 will be unique).

In addition to the stores, ANDS funded a number of pipes between metadata stores and enterprise systems, as well as metadata stores and data stores. Some of this activity occurred through the Collection Stores projects (in particular the Research Metadata Store Hub). As well as this, ANDS funded the development of a solution at Monash University to provide feeds of Activities and Parties information to ANDS from enterprise Oracle systems.

In late 2011, in a variation on the EoI approach used in Data Capture, ANDS offered Metadata Stores funding to 22 institutions to improve their existing metadata store infrastructure. Particular features of this program included:

1. Larger institutions were invited to participate, rather than all institutions (on the assumption that they had more research data to manage).
2. All institutions who were offered funding were offered the same amount (on the assumption that the costs to deploy were likely to be roughly equivalent for each institution).
3. Institutions that had already received Metadata Stores funding received an amount discounted by the existing funding – in one case this meant no funding went to one of the Data Capture institutions.
4. Funding could be used to enhance existing solutions, deploy new solutions, or provide improved connections to institutional context.
5. Institutions were expected to demonstrate a whole-of-institution commitment to their metadata infrastructure.
Institutions needed to install a metadata store solution that complied with a number of mandatory deliverables (in broad terms, a solution that was integrated with other data infrastructure and provided at least a feed to Research Data Australia) and which could also address a range of optional deliverables. All the resulting projects are therefore similar, and vary in detail or specific requirements, rather than overall direction. As such, the individual projects are not documented here. All but two of the institutions (Curtin University of Technology and University of Wollongong) who were offered this funding took it up and the vast bulk of the projects commenced in the first half of 2012.

The following institutions accepted Metadata Stores funding under these conditions:

- ANU
- CSIRO
- Deakin University
- Flinders University
- Griffith University
- James Cook University
- La Trobe University
- Macquarie University
- QUT
- RMIT
- Swinburne University of Technology
- University of Adelaide
- University of Melbourne
- University of Newcastle
- University of New South Wales
- University of Queensland
- University of South Australia
- University of Sydney
- University of Tasmania
- University of Technology, Sydney
- University of Western Australia

2.3.3 Activity/Deliverables for 2012-13

Of the 24 research institutions offered funding as described above, all proceeded to develop a Project Proposal. 22 went on to sign contracts. Both Curtin University (MS16) and the University of Wollongong (MS24) withdrew because the timing was not quite right. Curtin also cited a shortage of Developer
resources. Access to IT resources and a nation-wide shortage of Developers was, and remains, one of the critical obstacles to developing sustainable Metadata Stores projects.

One of the special challenges associated with this Program was that the parts did not come together to form a whole until completion. This meant there were few, if any, points of intervention for ANDS should a project lose its way. To mitigate the risk, all Project Plans were rigorously assessed for coherence. However, discontinuity of staff was another critical and unpredictable obstacle that served to delay the progress of a number of projects.

ANDS did not mandate any particular software solution. However, institutions were strongly encouraged to enter into consortia arrangements in order to maximise the value of the available funding. During 2012, there was a protracted period of software selection. By December 2012 all had committed to a software solution: 4 universities had chosen to deploy VIVO, 15 universities ReDBox/Mint, while UQ and the CSIRO built their own. ANU used the ORCA software (supporting Research Data Australia – RDA). UNSW built a hybrid system using Mint and their home-grown ResData.

Graph 1. Comparison of software choices and development environments (see Table 3)

Metadata Stores had more fixed deliverables than the more bespoke Data Capture and Seeding the Commons Programs. This proved to be very useful for the Project Managers when defending their deliverables against the agenda of Senior University staff seeking to channel project funds into pet projects.

From the outset, the most challenging Deliverable was D4: a demonstrated alignment of metadata records about Parties with the ARDC Party Infrastructure Project, with researcher descriptions contributed to the NLA, and with People Australia identifiers for researchers recorded against researchers.

Concerns were expressed about the robustness of the NLA system. In particular, the step involving a manual matching process inserted into what was mainly an automated process. The concerns were justified in as far as the NLA (like many universities) had severely constrained Developer resources to apply to a range of unexpected behaviours requiring work-arounds. See NLA/Trove unexpected behaviours: [http://metadata-stores.blogspot.com.au/2013/02/nlatrove-unexpected-behaviours-wed-13.html].

While multi-institutional consortia did not develop to the extent that ANDS had hoped, there were successful collaborations around several of the Optional Deliverables involving the ReDBox community. Notable, was the collaborative development of Data Management Planning and Reporting tools. However, these collaborations also caused dependencies and the consequent slippage of some projects due to having to wait for components to be completed by others.

The Metadata Store project team at the University of Newcastle, led by Vicki Picasso, provided a fine example of shared purpose and generosity of spirit. In addition to implementing a comprehensive solution to improve the management of research data at UoN they generously shared their experience and solution
design materials with other universities and made a major contribution to the development of a national community around the ReDBox software solution. One of the many success stories to come out of the Metadata Stores Program is the role of the ReDBox community in the Governance and development of the software.

Another important but largely untapped community of interest in the University sector is the community of Steering Committee members. People who are members of these Committees are people of influence and power. As Project Teams are wound up, these people could well become ANDS primary constituency. See Table #4 as well as Names of members of Steering Committees: [https://intranet.ands.org.au/ands-programs/metadata-stores-ms/process-documents/MS-SteeringCommittee-names-June2013.xlsx/view]

While the EIF funding rules required that the Metadata Stores Program had a focus on building software infrastructure, it is worth noting that when a project became effective as an institution-wide solution then it quickly becomes a change management project. Although institutional change was flagged by ANDS as being a critical component of the in-kind support of the project, the extent to which managing change quickly became the overriding concern is noteworthy. See LaTrobe video: http://www.youtube.com/watch?v=hmVxFObdk6Q

Tasmania (MS12), WA (MS17), Swinburne (MS26) and Macquarie (MS19) universities all had a difficult gestation due to a rapid turnover in Project Managers and finding adequate IT support. Due mainly to the hands-on support of the CLO, Luis Martinez Uribe, UTas was able to submit its Final Report before June 30th 2013.

A small number of projects had not completed by 30th June 2013; these were being actively managed.

2.3.4 Program Highlights, Issues and Breakthroughs

In order to facilitate community development a Metadata Stores blog was established to support the weekly gathering of people working on the projects using GoTo Meeting, aka Wednesday Data Surgeries. See [http://metadata-stores.blogspot.com.au/]. These gatherings proved very successful and were frequently filled to capacity (max 26). Success was not only because each week a member of the community would discuss some matter relating to their project but also because people got to know each other and understand the context of their projects in the sector. As a result, they soon began to trust and support each other. The development of this community may also be one of the significant unexpected benefits of the Program – that a community of interest has been developed strong enough to continue beyond the program. Some of the weekly presentations were recorded and can be found on the ANDS YouTube Channel or accessed through the Metadata Stores blog video tab. See [http://metadata-stores.blogspot.com.au/p/v.html]. Detailed instructions on the recording and permissions processes surrounding the conduct of the Data Surgeries have been documented on the ANDS Intranet.

A second highlight has been the growth of a community of practice and a community of development around the ReDBoX solution. As the solution implemented by the greatest number of institutions, the potential was always there for such a community, but it has come together in a very successful way. A ReDBoX Community 1.5 day event was held at Flinders University in February 2013, and the community has been successful at co-ordinating the development of new functionality.
An issue that all of the Metadata Stores projects have had to deal with has been the integration with the Trove NLA Identity Manager infrastructure. In particular, the task of creating and disambiguating Trove identities has proved more cumbersome than anticipated. If the Metadata Store projects were being commissioned now, this requirement would probably require ORCID integration instead, but a year ago the status and uptake of ORCID was less clear.

2.3.5 Program Learning

The funding program for metadata stores has been enthusiastically taken up by our partners. We believe that this is because the effort put in so far in other programs has moved the thinking in institutions forward so that they now see the value in having a metadata store (or, as we would now call it, an institutional research data asset registry). The lesson here is that sometimes one needs to wait for the right time (and/or work to make the time right).

2.4 Public Sector Data

2.4.1 Overview of program

Many areas of research are heavily dependent on government data – from cadastral data to economic data to government-organised surveys – or could increase their use of such data if it were more widely discoverable and accessible. The responsibilities inherent in data custody are a shared challenge and include the need to address preservation, access and description. As such there is a very close potential relationship between ANDS’ concerns and those government agencies that are custodians of data or that are influential in data policy.

The Public Sector Data program provides the infrastructure necessary to ensure that feeds of data collection descriptions are made available from a range of public sector agencies. Identified agencies include producers of research data, such as the Bureau of Meteorology (BOM), the Australian Bureau of Statistics (ABS), GeoScience Australia (GA), CSIRO, Departments of Primary Industry (DPI), Australian Urban Research Infrastructure Network (AURIN), Department Sustainability, Environment, Water, Population and Community (SEWPaC), Marine Community. Owners of data gathering activities and collections, such as the museum and library sectors, which might be possible inputs to other research activities, are also in scope.

ANDS also needs to maintain and develop stronger relationships with other organisations with significant data holdings or interest in these areas such as the National and State Archives Australia and the Australian Government Information Management Office (AGIMO), for example. Finally, ANDS explores ways to incorporate public data collected by citizens, through exemplar projects.

The key deliverable from this program is to make existing public sector data resources more discoverable to the research community and to work with federal, state and territory government agencies to improve access to data. Activities will vary across agencies according to their existing infrastructure and the types of data being made available. In all cases there will be a strong preference to have data services exposed using relevant international standards.
The Public Sector Data program was originally allocated a $10m budget in the ARDC Draft Project Plan. During the review phases for ANDS mid 2010 this budget was reduced to $6.45m. This was as a result of discussions with key government agencies in the first quarter of 2010 where they identified that their desire was for capability assisted infrastructure development from ANDS in preference to the provision of funding to undertake the infrastructure development themselves.

### 2.4.2 Outline of projects

ANDS has either entered into contracts for, or in negotiation for the following Public Sector Data projects and engagements:

<table>
<thead>
<tr>
<th>Agency or Institution or Project</th>
<th>Project Status and Description</th>
</tr>
</thead>
</table>
| EIF003 – Auscope SISS Deployment | Project complete. 103 earth science collections contributed to RDA. Potential for geology National Collection.                                                 |}
|                                 | **AusScope Discovery Portal**                                                                                                                                 |
|                                 | **Auscope in ANDS Registry**                                                                                                                                     |
| EIF024 – Australian Legal Information Institute (AustLII) | Complete. 501 collections contributed to RDA. Potential for further definition in National Collections.                                                        |}
|                                 | **AustLII**                                                                                                                                                     |
|                                 | **AustLII in ANDS Registry**                                                                                                                                       |
| EIF041 - PowerHouse Museum (CAMD/MA) | Complete. 1056 collections contributed to RDA. Software module for the creation of collections descriptions developed for major collection management software and released into open source for uptake by other vendors. Canvassing and initial contribution of university museum collections to MME. |}
|                                 | **Museum Metadata Exchange**                                                                                                                                      |
|                                 | **Museum Metadata Exchange in ANDS Registry**                                                                                                                     |
| National Criminal Justice Research Data Network | Complete. Data hosted on ADA. Discussions underway with ADA to obtain a metadata feed from their archive to RDA.                                               |}
|                                 | **NCIRDN portal**                                                                                                                                                 |
| GeoScience Australia | 19342 collections. Have added ANZSRC:FOR codes to collections. Work proceeding on minting DOIs (Digital Object Identifiers) via ANDS and adding citation metadata to nominated collections. |
|                                 | Have nominated certain National Collections to be hosted at National Computational Infrastructure (NCI).                                                        |}
<p>|                                 | <strong>Geoscience Australia in ANDS registry</strong>                                                                                                                          |</p>
<table>
<thead>
<tr>
<th>Organisation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian Institute of Health and Welfare</td>
<td>17 significant collections contributed to RDA. Intend to provide additional collections when resources are available. AIHW in ANDS registry</td>
</tr>
<tr>
<td>Australian Bureau of Statistics</td>
<td>Code developed for ABS to harvest time series records to RDA. ABS looking to secure funding to implement. Investigations to add ABS.stat as a ‘see also’ service in RDA underway. Participation in a demonstrator project to make census data available for AURIN. Plan to harvest collection, service, party and activity metadata via the AURIN project. AURIN will source from ABS and other data hubs.</td>
</tr>
<tr>
<td>Bureau of Meteorology</td>
<td>Engagement here is pending an enterprise reorganisation of their data and metadata. However there is potential in the next period to expose, via RDA, some of the urban water data they hold. Constructed a hierarchy of 14 collection records to describe the Australian Water Resources Assessment 2010. BOM in ANDS Registry</td>
</tr>
<tr>
<td>Murray Darling Basin Authority</td>
<td>Engagement on hold until MDBA can implement new data management processes. Will advise on readiness to re-engage. Consultation of data holdings and metadata framework underway.</td>
</tr>
<tr>
<td>Public Records Office of Victoria</td>
<td>Project complete - 14475 collections, 2496 parties and 136 activities. Software solution shared with State Records Authority of NSW who have hence contributed 15759 collections, 3789 parties. State Archives of Queensland intend to apply the same software solution. NAA have also expressed interest in providing records. Working with Council of Australasian Archives and Records Authorities to ultimately expose all Australian archives data. Public Record Office Victoria in RDA State Records Authority of NSW in RDA</td>
</tr>
<tr>
<td>Atlas of Living Australia</td>
<td>Project complete. Developed software that they then reconstructed and now maintain for providing collection and party records to RDA. 118 collections and 24 parties ingested. Data sets, providers and spatial layers included. They may proceed to describe species. ALA in ANDS Registry</td>
</tr>
<tr>
<td>Australian Antarctic Division</td>
<td>Project complete. 2136 collections and 848 parties provided. DOIs minted via ANDS and citation metadata included in nominated collections. AAD in ANDS Registry</td>
</tr>
</tbody>
</table>
2.4.3 Activity/Deliverables for 2012-13

During this period, all PSD contracted activities were successfully finalised.

Australian Institute of Health and Welfare (AIHW) were unable to commit resources to ANDS activity during the reporting period. ANDS has implemented technical changes to support AIHW metadata model and will be ready to engage, when AIHW have the required resources.

ANDS has met with Australian Government Information Management Office (AGIMO) several times during this reporting period. AGIMO have communicated the scope their collection strategy and will advise ANDS when the AGIMO infrastructure is stable.

ANDS is working with Australian Bureau of Statistics on the development of a data service using the new national geography (ASGC) through the provision of an on-line capability to manipulate the data, visualize the data and conduct spatial and statistical analysis and modelling of the data using a suite of e-research tools that already exist or are being developed through AURIN (Australian Urban Research Infrastructure Network) and will be incorporated within the AURIN e-infrastructure. The intent is to provide access not only to the ‘raw’ census data (i.e. the aggregated absolute data for specified levels of geography), but also to ‘value added’ data (ratio data such as percentages, and benchmarked indicators such as Location Quotients) which are typically used by researchers as operational variables in analysis and modelling and which can be generated through an automated process. This work is ongoing.

Bureau of Meteorology are undergoing an internal data management improvement program, ongoing into 2014. BOM will enable SISS in their Production Environment as part of this improvement program, enabling harvesting of metadata by ANDS. In the interim, ANDS partnered with BOM to deliver 14 Australian Water Resources Assessment records to Research Data Australia, as part of the Australian Urban Water Collection.

The Australian Ocean Data Network (AODN) and ANDS have developed a strong partnership. AODN has provided over 12000 marine science data collection descriptions to the ANDS Registry. AODN is an aggregator of data collected by the Australian Marine Community. Primary datasets are contributed by the Integrated Marine Observing System (IMOS) and the six Commonwealth Agencies with responsibilities in the Australian marine jurisdiction (Australian Antarctic Division, Australian Institute for Marine Science, Bureau of Meteorology, CSIRO, Geoscience Australia and the Royal Australian Navy). ANDS worked together with AODN to make significant improvements to the quality of the metadata being provided. ANDS also funded work for IMOS to expose marine datasets from South Australia Spencer Gulf Fisheries. This work was successfully completed, with 33 quality datasets described and harvested. The Final Report has been submitted and approved.

During ANDS’ partnership with the Australian Antarctic Data Centre (AADC), we took the Java software developed for Atlas of Living Australia and extended it. This enabled the transformation of AADC source metadata (available in ISO19115 format) into ANDS Registry-compliant metadata. 2136 Collection records
and 848 Party records have been harvested and are currently accessible via Research Data Australia. AADC has successfully minted production Digital Object Identifiers, and included their values within their source metadata. The .csv crosswalk for AADC has since been updated so that Citation Metadata is now constructed for the ANDS registry records; DOIs are now included where available.

The Murray Darling Basin Authority redirected resources to data management improvement projects. They have expressed interest in working with ANDS to expose their data holdings, once they have established processes.

During ANDS’ partnership with Atlas of Living Australia, Java application software was developed to transform web-accessible source metadata (available in JSON format) into ANDS Registry-compliant metadata. The software enables the description of datasets (236), data providers (49) and spatial layers (306). The transform component of the software can be configured within a crosswalk from the .csv format.

ANDS worked intensely with the Australian Institute of Marine Science (AIMS) during this reporting period. Initially, we developed metadata feeds from the e-Atlas data hub, then worked with AIMS to determine their data management requirements. From this activity, we re-established the AIMS metadata feed and will continue to work with AIMS to improve metadata and data capture. AIMS has provided 371 marine data collection descriptions to Research Data Australia.

Public Record Office of Victoria (PROV) completed a contracted project during this period. PROV lead the project on behalf of the Council of Australasian Archives and Records Authorities (CAARA). Government archives across Australia hold massive quantities of documents which are of great value to researchers, particularly in the humanities and social science fields. The project developed software tools for the automated extraction of metadata about these holdings and their transformation into the format required for delivery to the ANDS Registry. PROV have provided access to over 14000 Collections to date. The software tools and other products developed are open source and reusable by other National and State archives and agencies. State Records NSW have implemented components and have provided over 15000 Collections to the ANDS registry. Queensland State Archive has expressed interest in implementing the solution. There is potential for an Archives National Collection. Targeted archives are National Archives, State Records of South Australia and State Records of Western Australia, with whom relationships have been established.

Geoscience Australia dataset metadata has been successfully harvested from their Geonetwork metadata catalogue, into the ANDS registry. 19342 Collection records and 3330 related Party records can now be accessed via Research Data Australia. The Collection records have been enriched to include ANZSRC classification codes that have been mapped from the Geoscience Australia subject keywords. Geoscience Australia are in the process of constructing request XML to send to ANDS Web Services, to mint Digital Object Identifiers for their datasets; ANDS brokers the transactions between each partner (GA in this case) and DataCite. ANDS are working with GA to improve the transformation of GA’s Collection records into ANDS Registry-compliant records. Subsequently, records have been improved to contain high-quality citation metadata so that dataset can be accurately cited, including a DOI where one has been minted. Record citation metadata will be sent to Thomson Reuters Web Of Knowledge to be indexed within the Data.
Citation Index; doing so will allow metrics to be collated, and will render GA records accessible within a wider context.

Auscope completed all contracted obligations and in addition, implemented an instance of SISS at the National Computational Infrastructure (NCI). The super compute capability of the NCI will enable researchers to access, process and manipulate large data sets provided via SISS. The available data sets include Geoscience Australia’s back catalogue of Landsat imagery. Also developed by this project was a Virtual Exploration Geophysics Laboratory (VEGL), which is a workflow engine enabling researchers to process geophysical data. The workflow utilises the Amazon Cloud. During the project, SISS was implemented into 12 Government agencies, with a total of 103 earth science collections harvested to date. The Bureau of Meteorology implemented SISS into their development environment and will roll out to their Production Environment as part of an enterprise wide data management programme in 2014.

CSIRO Water Resources Observation Network project successfully developed a robust web interface for the CSIRO Data Access Portal. This web interface and the related data management infrastructure have enabled CSIRO to establish an enterprise wide data management and delivery system and service. The infrastructure includes an interface with Research Data Australia. During this reporting period, CSIRO has provided over 560 Collection descriptions, including from disciplines such as physical sciences, environmental sciences, earth sciences, biological sciences and agricultural and veterinary sciences.

Currently, the ANDS Registry contains a significant quantity of high-quality datasets that have been harvested from Geonetwork metadata catalogue instances managed by our partners including: AuScope; Australian Ocean Data Network; IMOS; Australian Institute of Marine Science and e-Atlas; Geoscience Australia; SMART Infrastructure Data Centre; TPAC CSIRO Statewide Hydrological DataSets; and UTAS-Institute for Marine and Antarctic Studies. The Geonetwork software component (an XSL Transformation file) that transforms metadata files into ANDS registry-compliant metadata has been improved (to varying degree) with customisation in some instances, by relevant ANDS’ partners. An XSL Transformation file exists per source metadata profile (ISO19115 Geographic Information, ANZLIC Profile, Marine Community Profile, etc.). ANDS has since applied effort in improving this transforming software component for the benefit of our partners. In particular, ANDS has consulted closely with Geoscience Australia while improving their XSL Transformation file that transforms from ISO19115 Geographic Information metadata. Consequently, GA’s XLS Transformation file has been improved such that the resulting ANDS registry-compliant metadata now includes citation metadata, including DOIs where provided; improved connections; and comprehensive rights information.

### 2.4.4 Program Highlights, Issues and Breakthroughs

PSD has continued its strong relationship with Geoscience Australia. During this reporting period, PSD assisted GA to develop an implementation strategy for Digital Object Identifiers. The strategy included change management for researchers and management, technical advice and development of business processes associated with data management. GA’s robust data management has resulted in over 19000 data collections being exposed via Research Data Australia (ANDS Registry). The data collections include GA’s back catalogue of satellite imagery, which has been made available via a collaboration between Auscope, GA and NCI (with the Auscope component funded by ANDS).
The ANDS-funded Government Archives Metadata Project was successfully completed, with at least 4 other archival agencies undertaking feasibility studies for implementing the solution. The project brought together state archivists and collections managers from State Archives and the National Archives to define requirements for the solution. The solution has resulted in the exposure of over 30000 state records for discovery by researchers, with more to come.

The Java application software that was developed for harvesting Atlas of Living Australia metadata (JSON format) and Australian Antarctic Data Centre metadata (ISO19115 Geographic Information) can be used to transform any ANDS’ partners’ web-accessible source metadata into ANDS Registry-compliant metadata; a crosswalk need merely be configured per use case.

The improvement that has been applied to the XSL Transformation file for Geoscience Australia’s Geonetwork metadata catalogue instance (for transforming ISO19115 Geographic Information Metadata to ANDS registry-compliant metadata) can be applied to all (current and future) ANDS partners whose metadata is web-accessible and in ISO19115 format. We will potentially be able to accept metadata in schemas other than ISO19115 now that we have gained experience in constructing and improving XSL Transformation files. This will fast-track our ability to harvest citation metadata, including DOIs where available; hence, enabling the provision of more, and higher-quality, citation metadata to the Thomson Reuters Web Of Knowledge Data Citation Index on behalf of our partners.

Data and information management is an issue that many government agencies are in the initial stages of addressing. PSD has engaged with agencies who, through our guidance, discover that they need to implement a more robust data management solution. Getting approval for these types of activities take significant time and undergo bureaucratic process, even when all levels of management agree on the benefits.

### 2.4.5 Program Learnings

The varying capability of the agencies’ ability to respond to requests for their data has made the progress of this program in increasing access to public sector data quite variable, requiring significant flexibility in approach. As the demand for public sector data increases, the policy drivers on open data become stronger, and acknowledgement of its use becomes evident and measurable the value and efficacy of this program is expected to increase.

### 2.5 ARDC Core Infrastructure

#### 2.5.1 Overview of program

Researchers and research organisations are increasingly expected to make their data public. This allows verification of research claims and the building of innovation upon previous work. Similarly, public funders of research are increasingly requiring public access to the inputs and outputs of research. Research assessment frameworks are also moving towards ways of acknowledging the publication of data as a research output.
The ARDC Core program is building out the national infrastructure services that enable research data to be published and discovered through a program of development of software utilities, both in-house and with partners, which enable organisations and individuals to publicise and share the collections they hold.

### 2.5.2 Outline of projects

**ANDS Collections Registry**

The ANDS Collections Registry enables the harvesting and depositing of description information about collections and the activities, parties and services (for example, RSS feeds) which relate to collections.

This information is exposed on the World Wide Web through *Research Data Australia*, a mesh of highly findable web pages. Research Data Australia is publicly available online at: [http://researchdata.ands.org.au/](http://researchdata.ands.org.au/)

The ANDS Collections Registry is a software application that holds and manages collection description information and associated context. Generally, ANDS participants set up an automated harvest of collection description information from their repository or data store to the ANDS Collections Registry. This enables the description information to be kept current via updates from the originating source. The ANDS Collections Registry supports a number of dynamic exchange and harvesting protocols to automate communication with managed data environments, including OAI-PMH and RESTful web services.

The ANDS Collections Registry accepts automated feeds of collection descriptions in an XML format called Registry Interchange Format – Collections and Services (RIF-CS). RIF-CS is based on the ISO 2146:2010 (Registry Services for Libraries and Related Organisations) standard. ISO 2146 is an international standard intended to operate as a framework for building registry services for libraries and related organisations.
Research Data Australia

Research Data Australia is a set of web pages (see Figure 2 for a sample) describing collections produced by or relevant to Australian researchers. It is designed to promote visibility of research data collections in search discovery engines such as Google and Yahoo, and to encourage their reuse. It is the face of the ANDS Collections Registry service. Research Data Australia includes discovery tools, spatial coverage display and numerous features that capitalise on the mesh of data provided through national and global partner services.
to improve quality and richness in the display and discoverability of content. ANDS has implemented a number of significant enhancements to Research Data Australia in 2012-13 (more details in Section 10.3.4) and will continue to enhance it in 2013-14.

Register My Data

Register My Data allows individuals to manually register a collection description through an easy to use online web form. Collection descriptions are stored in the ANDS Collections Registry and are publicly discoverable through Research Data Australia. To be registered, collections must be accessible online.

Register My Data is an entry level product which requires a minimum of researcher effort to use. Collection descriptions must include a title, URL and a short description of the collection. Researchers may also add additional information about other contributors, subject keywords, spatial coverage and access rights.

ANDS prefers to harvest collection description information automatically, at the institutional level, as this allows for the responsibility of ongoing maintenance of collection description information to rest with the institution. However, this is not always possible. Register My Data is intended for use by researchers at organisations where there is no formal data archiving service and where ANDS has no distributed services in place. Effectively an institutional Register My Data is available to research organisations that use the machine to machine interfaces of both the ANDS Collection Registry and Identify My Data.

Identify My Data

Identify My Data provides cost free creation (called minting), resolution and management support for persistent identifiers for the Australian research and cultural collections sectors.

A persistent identifier (PID) is a number or a code that is allocated to an object or a resource as a long lived identifier. Persistent identifiers provide a globally unique identification. Persistent identifiers can be used to create hyperlinks on a web page. When clicked on, the hyperlink will take the web-user to the URL associated with the identifier.

Persistent identifiers are useful because when an object or resource changes location on the internet, the persistent identifier does not change. The owner of the resource is able to communicate with the persistent identifier database and change the object location associated with the persistent identifier. All links which use the persistent identifier will then point to the new location without the resource owner needing to update them.

Identify My Data can be from either from a web form, or via machine-to-machine transactions. Identify My Data is useful for:

- software developers needing to build software which assigns persistent identifiers to objects within their application where there is a requirement to assign, update and resolve identifiers.
- individuals wishing to provide persistence for citing materials held on their web site who may want to assign identifiers to individual pages, papers, or a web site
- institutions acting as an authoritative source of information about people, organisations or community vocabularies where globally unique identifiers need to be assigned to individual records
institutional repositories not wishing to manage their own identifier service who may wish to assign persistent identifiers to their holdings.

Identify My Data uses the international Handle System developed by the Corporation for National Research Initiatives (CNRI). The Handle System provides identifier resolution services for use on the Internet. It includes a set of protocols that enable a distributed computer system to store identifiers – known as handles – and to resolve those handles into the information necessary to locate, access, contact, authenticate, or otherwise make use of the associated resources. This allows the handle of an item to persist despite changes of location and other changes. The Handle system is very robust and is widely used internationally among repositories.

**DOI Service – Cite My Data**

DOIs form the basis of publication identification in current academic publishing practice. ANDS' aim is to treat 'published' datasets in a way analogous to journals and monographs. Citation indexes already use DOI as the basis of counting citations, and discussions with ThomsonReuters and Scopus have confirmed their willingness to deal with DOIs allocated to datasets. DOIs are known and used by researchers, publishers and citation indexes.

Research communities understand DOIs because of their use in journal publications. The experience here and overseas is that data creators and managers are more comfortable using the same persistent identifier infrastructure for published data sets as they have for journal articles. ALA, APN, CSIRO Land and Water, AANRO, ANUSF, and ASSDA have all confirmed this in informal consultation.

ANDS has joined the Datacite international consortium. The goal of this consortium is to establish a not-for-profit agency that enables organisations to register research datasets and assign persistent identifiers to them, so that research datasets can be handled as independent, citable, unique scientific objects. DataCite is a coalition of nationally-focused organisations dedicated to registering and allocating identifiers to scientific datasets.

ANDS has developed minting software for Datacite, and contributed to the development of a minimum metadata set. During 2012, ANDS has made significant enhancements to the pilot Cite My Data service (based on the feedback from our pilot partners), consequently launching it as a full production service.

This ability for Australian eresearch organisations and data centres to allocate citation optimised identifiers (DOI) is one of the enabling services which will allow data collections to be cited and tracked as first-class objects in the scholarly communications cycle. The maintenance and improvement of this infrastructure (to continue to meet the requirements of eresearch organisations) will continue in 2013-14.

**Vocabulary Service Project – Classify My Data**

To improve the accuracy and precision of research reporting, most research communities use fixed terminologies, concepts, units of measure and so on. Some, although not all, of these items take the form of controlled vocabularies. Given the importance of controlled vocabularies to science, a national Vocabulary service would benefit the research community by providing a managed infrastructure and support service that would not only make it easier to access and use existing vocabularies and but also provide new
opportunities for other communities to formalise their terminologies and bring it under a common framework for their sustainable management and governance.

Controlled vocabularies are widely used in library and information sciences to better organise and describe knowledge. Controlled vocabularies help standardise the use of language in bibliographic and metadata descriptions and enhance the precision of retrieval against known terms and concepts.

In order to trial concepts and gather requirements from end users in the research community, ANDS has designed a number of prototype components of the Classify My Data vocabulary service which provide a set of web and online services to support the creation, management, and publication of human and machine-readable controlled vocabularies for use by the Australian research and higher education sector.

Among many other benefits, the service will provide the research community with a managed infrastructure and support service that will not only make it easier to access and use existing vocabularies, but also provide new opportunities for other communities to formalise their terminologies and bring it under a common framework for their sustainable management and governance.

In 2012-13, development of the first phase of this project began, with the following functions being designed and piloted:

- Upload vocabularies (GUI and web services)
- Retrieve vocabularies (via web services)
- Update vocabularies (GUI and web services)
- Retire vocabularies
- Download vocabularies
- Browse & search vocabularies
- Search vocabularies
- Enabling integration of vocabularies into existing systems

Work in 2013-14 will continue to build on this initial phase, incorporating feedback from the e-research community, in order to improve and launch a full production ANDS service.

ANDS convened a vocabulary services technical working group with input from NCRIS facilities and commonwealth agencies.

**Party Infrastructure Project**

The objective of this project is to improve the discovery of research data and research publications by linking them through common researchers and research groups. ANDS has partnered with the National Library of Australia to provide infrastructure to allow Australian researchers and research organisations to more easily use an independent party identifier when publishing their information about data collections. Figure 4 shows the workflows for content providers using the NLA Party Infrastructure.
ANDS has also become a member organisation of ORCID, an international body offering globally unique identifiers to identify researchers. Through this partnership, ANDS has been able to improve the capture and identification of researchers within the ANDS Collections Registry, as well as promoting the research community’s adoption of ORCID and linking of their collections in the ANDS Collections Registry with their ORCID profile. See diagram below:
Having a persistent identifier assigned to researchers enables different institutional systems that manage information about researchers to interoperate. The institutional repository, research management systems and other systems providing information about researchers can exchange information using a common identifier. That identifier can also be linked to other information external to the institution, such as former and concurrent affiliations and publications as well as cross-institutional collaborations.

**Location Infrastructure Project**

An important goal of the Australian Research Data Commons is to enable cross-disciplinary discovery of related research data, and spatial location is a vital linkage mechanism in this process. The value of the data commons will be increased if the dataset descriptions include spatial coverage data encoded as geographical points or polygons rather than just text. ANDS’ vision for a data commons would see non-GIS-experts from arts, humanities, and science able to enrich their discipline specific data with standardised spatial information.

Achieving this goal requires the establishment of a robust national infrastructure that would allow place names to be validated by both individuals and software systems against an Australian gazetteer service in an efficient manner. There will need to be distributed sources of gazetteer data, depending on jurisdiction, feature types, temporal coverage and language. A comprehensive national service will need to provide interoperable query services across these sources, irrespective of their differing construction.
This infrastructure is intended to increase the amount and quality of spatially-marked-up research data. This will enable new kinds of research and innovation based on new data linkage and data merging opportunities. The infrastructure aims to unlock significant innovation and productivity. It will bring benefits well beyond the research and innovations sector.

ANDS has partnered with Geoscience Australia (GA) to develop and run an online Gazetteer location service. This project provides a publicly available interface (both web and machine readable) to an authoritative national gazetteer of place names and other useful spatial information. GA is the national geoscience research and spatial information agency.

The national value of this project lies in its ability to enable more spatially enabled data to be produced by any researcher, group, or data repository. GA will develop, deliver and run the Australia Gazetteer service as a national service into the future.

The specific deliverables include:

- Development of a gazetteer data schema, database, query and web service
- Development of a user interface with search and display mapping functionality
- Commissioning systems and services

Data from this service is publicly available at no cost (previously these data were subject to charge).

**Activity Infrastructure Project**

The project will leverage the significant information holdings of the Australian Research Council (ARC) and the National Health and Medical Research Council (NH&MRC) with regard to research activity underway in Australia.

Proposals for projects with the ARC and NH&MRC are currently under development for them to maintain persistent URI identifiers for all of their awarded research grants which will resolve to structured data which can be accessed by other systems as well as by humans. However, this is a longer-term goal which is not likely to reach final implementation until later in the ANDS project.

In the interim, these records have been ingested as Activity records into the ANDS Registry, and are available for linking to collection records by ANDS participants, which will increase and add value to the mesh of navigable research records discoverable through Research Data Australia.

### 2.5.3 Activity/Deliverables for 2012-13

- A key activity for the year 2013 has been the substantial optimisations made to the software backbone of ANDS utility services, including a complete overhaul and modernisation of the codebase underpinning the ANDS Collections Registry and Research Data Australia. These upgrades facilitated significant performance improvements, a large number of feature requests and laid the foundation for more rapid development and maintenance of these services into the future. Further, the new codebase has been optimised for open source contribution and reuse, enabling simpler adoption of ANDS software by other eResearch organisations. All ANDS software is open sourced and the most recent release is available at: [http://github.com/au-research/](http://github.com/au-research/).
The ANDS Collections Registry was enhanced to support harvesting from providers that are not able to provide data the registry’s own RIF-CS format. A further outcome of the software overhaul was the ability to support multiple database storage engines, facilitating adoption of ANDS software in a wider variety of environments, independent of ANDS’s preferred operating system and database system.

The ANDS Harvester (the software responsible for regularly collecting ("harvesting") metadata from partner organisations and depositing the most up-to-date descriptions in the Registry) was upgraded to optimise support for commonly used metadata store software and improve performance of large-scale providers. Notably, the project delivered a 90% reduction in delay when harvesting from ANDS’s largest partner organisations.

Many of the advanced data capture capabilities in the Register My Data tool have been packaged and published in a collection of “software widgets” which partner organisations can easily reuse in their internal data registration workflows. With an emphasis on a low barrier to entry, these widgets and supporting service infrastructure provide “plug and play” compatibility with a number of ANDS services, aiming to facilitate institutions improving the richness of collection description information captured at the source.

ANDS commissioned a review of the graphic design of Research Data Australia and implemented significant updates to modernise the appearance of the web pages in RDA. Improvements were also made to the navigation and search engine optimisation of the site, with enhancements in the display and clarity of information presented. New functionality was added to enable the highlighting of new, unique and interesting data collections through the “Spotlight on Research Data” tool and specialised visualisations were added to assist in the navigation of collections with extensive contextual information. ANDS also provided a new mechanism whereby eresearch organisations are able to publish “institutional pages” in Research Data Australia as a method to showcase all of the collections and contextual information (such as grants and researchers) which are associated with that research organisation.

ANDS launched two new service components as part of the prototype Classify My Data service. The ANDS Vocabulary Catalogue allows eresearch organisations to register their controlled vocabularies, providing services such as description listings, browsing capabilities and download & versioning functionality. Paired with the catalogue, the ANDS Vocabulary Service provides machine-to-machine descriptions of controlled vocabularies using the Resource Description Framework and a best of breed “linked data” approach.

ANDS has used both the Vocabulary Catalogue and Vocabulary Service to, for the first time, publish the RIF-CS controlled vocabularies in a formal and structured format. Further, in consultation with the Australian Bureau of Statistics, ANDS supported the publishing of the Australia New Zealand Standard Research Classifications (ANZSRC) using the Classify My Data prototype. As a demonstration use case, the ANDS Vocabulary Service has been implemented to be used internally within the “Browse by Research Area” tool in Research Data Australia and to provide the vocabulary fields in the Register My Data interfaces.

Many of the advanced data capture capabilities in the Register My Data tool have been packaged and published in a collection of “software widgets” which partner organisations can easily reuse in
their internal data registration workflows. With an emphasis on a low barrier to entry, these widgets and supporting service infrastructure provide “plug and play” compatibility with a number of ANDS services, aiming to facilitate institutions improving the richness of collection description information captured at the source.

- After consultation with a number of partners in the pilot program, ANDS launched the Cite My Data service as a full production service, with a number of enhancements to the web services available to partner institutions. Further, a user-friendly web interface was established to assist implementers at partner institutions to setup and test their integration with the ANDS DOI service.
- Location Infrastructure: the Gazetteer project was launched into production by GA in the previous reporting period. ANDS has developed a software widget to facilitate integration of this infrastructure into data archives.
- Party Infrastructure: Major progress has been made towards integration of ANDS infrastructure with the global ORCID researcher identification infrastructure. The NLA are operating the national people identifier system and ANDS has provided significant technical expertise to support institution integration.
- ANDS is a formal partner in the EU-funded ODIN project to explore linkage of identifiers for researchers (ORCID) and datasets (DataCite).
- The service desk, change management system, and reporting functionality continue to be an essential part of the operation of a robust national services and infrastructure. Improved business intelligence reporting mechanisms have been added to provide an analytical view of ANDS services.
- The RIF-CS Advisory Board, constituted and independently chaired by partners and members of the community, continues to give advice to ANDS on changes to the registry interchange format. The Board reviewed and approved a new version of RIF-CS v1.4 in October 2012 and has met regularly during this year to deliberate on a new version for October 2013. ANDS has made considerable contribution to this group, establishing a direction for catering for the linking of collection descriptions with publications and improving the service capabilities for accessing data.

On the following pages are the service usage reports for 2012-13 with historic figures from the previous years:

<table>
<thead>
<tr>
<th>PUBLIC SYSTEM</th>
<th>Handle Service:</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handles Minted</td>
<td>4,279</td>
<td>6,323</td>
<td>6,830</td>
<td>9,015</td>
<td></td>
</tr>
<tr>
<td>Trusted SW Agreement</td>
<td>10</td>
<td>11</td>
<td>14</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td><strong>DOI Service (Cite My Data):</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>DOIs Minted (production)</td>
<td>-</td>
<td>-</td>
<td>29</td>
<td>1,712</td>
<td></td>
</tr>
<tr>
<td>Registered Publisher Agents</td>
<td>-</td>
<td>-</td>
<td>12</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td><strong>Registry:</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Data Source Admins</td>
<td>-</td>
<td>46</td>
<td>185</td>
<td>224</td>
<td></td>
</tr>
<tr>
<td>Provider Org</td>
<td>12</td>
<td>21</td>
<td>59</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Total data source feeds</td>
<td>12</td>
<td>45</td>
<td>122</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>11</td>
<td>30</td>
<td>60</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Automatic</td>
<td>1</td>
<td>15</td>
<td>62</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Publish my Data</td>
<td>51</td>
<td>63</td>
<td>60</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3: Service usage reports for 2010-13

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Data Australia:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Page Views</td>
<td>2,906</td>
<td>26,746</td>
<td>72,990</td>
</tr>
<tr>
<td>Page Hits</td>
<td>1,319</td>
<td>23,899</td>
<td>39,957</td>
</tr>
<tr>
<td>Filtered Page Views</td>
<td>894</td>
<td>1,960</td>
<td>5,473</td>
</tr>
<tr>
<td><strong>TRIAL SYSTEM:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handle Service (Test):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handles minted</td>
<td>3</td>
<td>30</td>
<td>84</td>
</tr>
<tr>
<td>Trusted SW Agreement</td>
<td>660</td>
<td>857</td>
<td>27,476</td>
</tr>
<tr>
<td><strong>Sandbox:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Source Admins</td>
<td>311</td>
<td>1,960</td>
<td>5,473</td>
</tr>
<tr>
<td>Provider Org</td>
<td>307</td>
<td>23,899</td>
<td>39,957</td>
</tr>
<tr>
<td>Total Records</td>
<td>2,392</td>
<td>26,270</td>
<td>7,519</td>
</tr>
<tr>
<td>Collections</td>
<td>3,753</td>
<td>32,416</td>
<td>40,233</td>
</tr>
<tr>
<td>Parties</td>
<td>29</td>
<td>44</td>
<td>114</td>
</tr>
<tr>
<td>Services</td>
<td>316</td>
<td>30</td>
<td>84</td>
</tr>
<tr>
<td>Activities</td>
<td>690</td>
<td>857</td>
<td>27,476</td>
</tr>
</tbody>
</table>

**NOTES:**

1. The DOI Service, Cite My Data, was launched last December 2011. This figure shows DOIs minted since its release.
2. “Page Views” are all Research Data Australia page views including hits from crawlers and robots (this is the raw data gathered from the web server logs). This method was used in 2009-10. The Google Analytics tool was used to gather page view reports from July 2010 to June 2012.
3. “Page Hits” are Research Data Australia page views cleansed by the Google Analytics tool (thus excluding robots, crawlers etc.).
4. “Filtered Page Views” are Research Data Australia unique page views from Google Analytics tool (filtering out repeat viewings by the same individual in the same session).
5. The numbers shown for the total records in Sandbox (trial system) for 2012 are only up to April 2012. The Sandbox environment was decommissioned on 30 April 2012 and records have been migrated to the Production (public system).
Figure 4: Handles minted

Table: ANDS Handle Service Report - Handles Minted

<table>
<thead>
<tr>
<th>Total</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUBLIC SYSTEM</td>
<td>4,279</td>
<td>6,323</td>
<td>6,830</td>
<td>9,015</td>
<td></td>
</tr>
<tr>
<td>TRIAL SYSTEM 5</td>
<td>-</td>
<td>15,399</td>
<td>32,937</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: User Agreements

Table: ANDS Handle Service Report - Trusted SW Agreement

<table>
<thead>
<tr>
<th>Total</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUBLIC SYSTEM</td>
<td>10</td>
<td>11</td>
<td>14</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>TRIAL SYSTEM 5</td>
<td>-</td>
<td>30</td>
<td>55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: “Trusted SW Agreements” are agreements for e-research organisations to access ANDS identifier services using the machine to machine interface.
Figure 6: Provider Organisations

Figure 7: Data Provider Accounts

Note: Data Provider Accounts are administrator representatives from organisations who manage feeds of information from their organisation to the ANDS registry.
Note: Page views in 2010 were simply unfiltered logs; from 2011 a filtered page view and unique hit reporting approach was adopted using standardised Google Analytics.

2.5.4 Program Highlights, Issues and Breakthroughs

The substantial upgrades to the software codebase have been a major highlight of 2013, with significant benefits being derived from better performing and more easily maintainable infrastructure underpinning the Australian Research Data Commons. The upgrades to ANDS software systems have meant that the system is scaling nicely to cope with growth. During this reporting year the ANDS registry passed the 100K record
mark. The migration to NeCTAR application hosting has supported this growth and ensured a consistent and stable hosting environment for ANDS’s national infrastructure services.

The maturation of the Cite My Data service has been a major focus for ANDS during this year and the system has continued to be improved, owing to the close relationships which ANDS maintains with its partner organisations. Demand for the service has been high despite the long-term nature of the commitment and rewards. Twenty five Australian research organisations integrating this infrastructure into their data archive services. Australian research organisations seem to see the potentially significant ongoing benefits of a global system to link researchers with their data output enabling acknowledgement and potentially reward.

ANDS has continued to adapt the direction of its software development, with a design imperative of supporting reuse of the ANDS software systems by other e.research institutions. An example of this software being a specific output of the project has been the implementation of a number of ANDS’s key feature additions being done in such way that these features can be easily packaged as “widgets” and purposed for reuse in the workflows and systems of other users in the e.research community. These widgets have enabled “drop in” components for capturing spatial information (using the Location Infrastructure – Gazetteer of Australia) as well as controlled vocabulary terms (using the ANDS Vocabulary service prototype). By promoting this methodology, we see the adoption of richer information capture tools as a key part of developing a cohesive and rich data commons.

The previous establishment of an ANDS service desk with a service/change management framework, system and processes has meant that service provision has also scaled nicely with the increase in the customer base.

### 2.5.5 Program Learning

ANDS has taken a deliberate and considered approach to the design and implementation of new infrastructure services in order to establish the requirements of the e.research community and deliver services which appropriately cater for the scale and interdisciplinary nature of the Australian Research Data Commons. Identifier services (such as Cite My Data) have begun to realise the value of this approach, with very high interest subsequent to the production release of the service, with feedback from the pilot being effectively leveraged in the implementation of the production system.

The ANDS “wholesale” approach to infrastructure services (taking an institutional-based engagement approach) has continued to bear fruit with a number of very large providers continuing to feed hundreds (and even thousands) of data collection descriptions into the ANDS Collections Registry and Research Data Australia. ANDS has moved quickly to cater for this rapid growth in service uptake, with considerable planning and execution of performance and usability improvements having taken place. This vigilance remains essential as usage of the services continue to develop.

ANDS has also benefited from the overhaul and carefully considered redesign of the software codebase this year, which has positioned the development team to take a more iterative and engaged approach to feature and service development. Further, with a number of ANDS software components now being packaged into reusable “widgets”, it is hoped that the richness of information being provided to ANDS will continue to improve, as e.research organisations are able to easily reuse these infrastructure components with a very low barrier to entry.
ANDS online services are now characterised by strong infrastructure partnerships and integration with global services including ORCID, DataCite, NLA, ARC, NHMRC, GA, and international commercial information services such as ProQuest and Thomson Reuters.

## 2.6 ARDC Applications

### 2.6.1 Overview of program

The aim of this program was to develop a range of compelling demonstrations of the overall value of the ARDC by bringing together a range of data sources combined with new integration and synthesis tools to enable new research or generate new policy outcomes.

Applications is leveraging the outputs from the other ANDS programs, which have been designed to:

- provide underpinning infrastructure to support discovery and citation (ARDC Core, in collaboration with International Infrastructure);
- enable rich metadata about data to be managed and accessible (Metadata Stores);
- make new data and associated metadata available from a range of instruments (Data Capture);
- make a selection of existing data and associated metadata available from the bulk of Australia’s research-producing universities (Seeding the Commons);
- make data and associated metadata available from government departments (Public Sector Data);
- work with RDSI to ensure a rich landscape of well-managed and reusable national data collections (National Collections); and
- provide the overall policy and practice frameworks to support better data management and reuse (Frameworks and Capabilities).

ANDS has been funded to bring about an Australian Research Data Commons. This has required a set of coordinated programs of activity that are described elsewhere in this Business Plan. The resulting infrastructure supports data discovery and access. Once accessed the data can be reused as is, but bringing together different data sets can enable new kinds of research. Before this can occur the data may need to be transformed or recoded. Once combined special analysis techniques may be needed to provide the right starting point for further research. There are many possibilities here across a whole range of research problems, and so ANDS is selecting a subset to demonstrate what is possible.

The output of the Applications program is a series of compelling demonstrations of the value of having data available for reuse. These should:

- result in data being transformed or integrated across multiple sources to produce new forms of information that enable innovative, high-quality research outcomes;
- deliver value to a high-profile research champion;
- be relevant to a range of government portfolios; and
- engage with national research capabilities.
2.6.2 Outline of projects

Consistent with these criteria for demonstrations of value, ANDS has funded a portfolio of 25 projects across two broad thematic areas: Climate Change Adaptation, and Characterisation of Biological Systems. There is also a collection of other projects to provide balance and demonstrate what is possible across a range of discipline areas. These projects have been carefully balanced across major research institutions, States/Territories, NCRIS Capabilities and national research priorities.

2.6.3 Activity/Deliverables for 2011-2012

The following Climate Change Adaptation projects have been funded:

- Tropical Data Hub (tropicaldatahub.org);
- Tropical Data Hub - Tools Development;
- Species Distribution Records (spatial ecology.jcu.edu.au/Edgar/);
- Climate Change Impacts Downscaling tool;
- Marine Video;
- Health Impacts of Climate Change;
- Climate Change Adaptation Information Hub; and
- Impacts on Ports Infrastructure.

The following Characterisation of Biological Systems projects have been funded:

- BPA X-omics tool;
- Data-Models-Papers interconnections (presented already at eRA 2011);
- Soils to Satellites (blogging at soils2satellites.blogspot.com.au);
- Mouse Brain Map/Tissuestack (early prototype at caivm1.qern.qcif.edu.au/desktop.html);
- Multiscale Kidney Imaging
- Genomics Virtual Laboratory Project;
- ALA Bird Distribution (the back-end services for the species distribution records project); and
- Human Chromosome 7 Proteome Browser (blogging at sites.google.com/a/ozhupohpp7.com/thehumanproteomebrowser/project-updates).

The following other projects have been funded:

- Public Open Space modelling;
- SMART Infrastructure dashboard;
- Founders and Survivors;
- CODCD/BMRI demonstrator;
- With the support from ANDS, the Chromosome 7 working group is the first to develop a viable approach to making the data accessible to scientists. “Our hope would be that all the other chromosome researchers will look at this browser and think, ‘We can use this for our chromosome.’ In fact it could be used as a uniform tool across this initiative internationally,” Prof Ian Smith says. “And there’s no reason why it could not be easily adapted to other fields and organisms.”
- Marine Virtual Laboratory Information System (blogging at marvlis.blogspot.com.au);
- Primary Production in Space and Time (blogging at episat-software.blogspot.com.au);
- TERN/ACEAS Project;
- Urban Climate Modelling; and
- Catchment to Coast.

The Applications program has started to deliver demonstrations of value over the reporting period, which is very pleasing to see. However, about half of the projects did not complete by the end of June 2013. In some cases, this is because the projects ran late. In others, this is because they submitted their final reports before the end of June, but our assessment (and possible reworking of the final reports by the projects in response) ran into the next reporting period. A number of the Applications projects had very aggressive timelines, so this is not entirely surprising. While this slippage is not ideal, ANDS decided to permit time extensions where to do otherwise would jeopardise the delivery of high-quality demonstrations of value.

Most of the projects decided to blog about their activities, and the results can be viewed at http://andsapps.blogspot.com.au/p/project-feed.html. A number of projects also produced videos. For examples, please consult:

- Edgar: http://www.youtube.com/watch?v=g8z2yFDYwHI, http://www.youtube.com/watch?v=sl6lEerfp0

2.6.4 Program Highlights, Issues and Breakthroughs

The main highlight of the Applications program has been the innovative ways in which different kinds of data are being brought together to answer new problems, and the evident enthusiasm of the researchers for what this makes possible.

Other highlights have been the appetite within the research sector for this kind of infrastructure activity, and the way in which a number of the projects in the Climate Change Adaptation space have strong connections (both personal and technological) with each other.

One of the issues confronting the Applications program has been the length of time taken to complete the entire process from commissioning to completion each project. The process from identifying a candidate through a series of discussions to describe the engagement through to contracting has taken up to 12 months in some cases. Once the projects commence, a number of them have encountered delays associated with changes in personnel and/or challenges in developing software with the required functionality. Despite this, all those that have completed have produced excellent results.

2.6.5 Program Learning

Engaging in discussions with research champions about barriers they face to bringing data together has been very instructive. A number of them commented that the kind of funding that ANDS has provided has been an excellent complement to national competitive grant funding. Competitive research grants can fund the
actual research but not its translation into software that lets researchers build on their insights. The ANDS Applications funding can fund the latter but not the former. Together, a rich range of possibilities becomes available. It would be good to identify a way to continue this sort of activity, perhaps in conjunction with something like the NeCTAR eResearch Tools and Virtual Laboratories programs.

2.7 International Infrastructure

2.7.1 Overview of program
The aim of this program is to work collaboratively with international organisations and partners to establish an effective international data-sharing environment for Australian researchers.

ANDS has been established to provide services to, and build infrastructure with, the Australian research and innovation sector. As described elsewhere in this document, ANDS is partnering with Australian research organisations, government agencies, and cultural organisations to build the Australian research data commons so as to provide a research advantage to Australian researchers.

Despite this essentially Australian focus for the implementation and outcomes of ANDS, there are still strong motivations for ANDS to engage internationally. Research itself is an international activity. Research organisations in Australia are in a global business and are involved in international collaboration and competition. Research infrastructure is also increasingly international.

ANDS will therefore pursue international opportunities in order to:

- join mature international infrastructure initiatives to provide better infrastructure and services at home;
- be exposed to alternative national, regional, and domain specific infrastructure implementation approaches to inform the ANDS approach;
- collaborate on and explore directions in international infrastructure initiatives with peer organisations to maintain our position of international prominence by pulling our weight; and
- influence developing international infrastructure initiatives to support the ANDS mission by adopting technologies and approaches compatible with those adopted by ANDS.

The intention of ANDS’ international engagements is to further its mission and take a leading place in emerging international infrastructure initiatives. The value of this international engagement on research data infrastructure includes:

- ensuring that Australian researchers use infrastructure that makes international engagement simpler;
- ensuring Australian researchers work with data that can be easily shared;
- ensuring that Australian researchers have an advantage in data cooperation;
- increasing the value of Australian research data investments by future proofing them;
- decreasing the cost of Australian research data investments by sharing the costs with international partners; and
demonstrating to stakeholders the value of the Australian approach to research data infrastructure by establishing international pre-eminence.

Key goals of ANDS in all of this activity continue to be emphasizing the importance of richly described and connected research data collections, and focusing on the significance of research institutions in managing research data.

2.7.2 Outline of projects

This activity has not involved the funding of any projects.

2.7.3 Activity/Deliverables for 2012-2013

ANDS actively participated with a range of counterparts to establish global research data infrastructure. This work will enable Australian researchers to be the beneficiary of this infrastructure and will also enable the distinctive elements of the ANDS approach to research infrastructure, most particularly the generalist collection approach and the strong institutional role, to influence the form of the global approach. Over the course of 2012-13, the main focus of activity was as the designated Australian Non-Government Sector organisation (NGS) for the DataWeb Forum initiative (later rebadged as the Research Data Alliance); this is being led by the NSF and NIST in the US and involves the EC as partners. The Alliance seeks to combine the bottom-up strengths of the Internet Engineering Task Force approach with a community-based governance model to work on data interoperability issues. The Alliance was formally launched at an event in Gothenburg in March. Ross Wilkinson was appointed as a member of the initial Council for the Alliance.

Beyond this, ANDS staff took place in a range of activities:

- ANDS is a partner in the EU FP7-funded ODIN (ORCID and DataCite Interoperability Network) project. This is working on connections between data, authors and publications. Other partners include DataCite, arXiv at Cornell, the British Library, CERN, Dryad at Duke University and ORCID Europe.
- ANDS ran the data plenary at the Second EU-AU Research Infrastructure workshop in Brussels in July 2012, as well as a collocated one-day data infrastructure workshop to develop plans for concrete data bridges in the Marine and Linguistic domains.
- ANDS continued to work directly with national counterparts in the UK, Netherlands, Finland, Germany and Denmark.
- ANDS also further developed its relationships with two NSF DataNet projects: the Data Conservancy and DataONE, as well as having Adrian Burton serve on the international advisory board of the whole DataNet program.

2.7.4 Program Highlights, Issues and Breakthroughs

The main highlight was the launch of the Research Data Alliance (RD-A). ANDS was designated the Australian NGS for the Research Data Alliance, and worked very hard to ensure the success of the establishment of RD-A and its launch. This initiative has the potential to bring about significant improvements in co-ordination of data interoperability – ANDS now needs to continue to help it succeed in so doing.
The main issue across all of the international activity is the tyranny of distance and time zone. It simply isn’t possible to have the same quality of interaction with international colleagues over voice or video linkages as face to face. This means that this area of ANDS will continue to require overseas travel, although this will be undertaken only when required and for significant events.

2.7.5 Program Learning

The most significant learning is the very high regard with which the Australian approach to research data infrastructure in general and ANDS in particular are held overseas.

2.8 Project Office

2.8.1 Overview of program

This program is designed to ensure the effective and efficient delivery of the ANDS-funded projects with proper reporting of outcomes. It has also adapted to support the delivery of all ANDS outcomes with the goal of delivering operational excellence.

2.8.2 Outline of projects

The function of this program has further evolved as ANDS has matured. It has gone from the task of development of ARDC infrastructure, through the management of a significant number of contracts as ANDS funded data management activities across the research and public sector to accounting for and communicating the outcomes of these activities.

Taking ‘Excellence in operational management’ as its goal, the program has worked on streamlining processes and increasing the cohesion of the ANDS group for more effective delivery of service by introducing standard practice, shared tools and cooperative processes.

This twelve month period within ANDS has been one of significant change and consolidation. The period has seen a number of people leave the team as well as the need to work through a planned reduction in full time staff. The main activity that has been undertaken to maintain the focus of operational excellence has been around maintaining the existing processes and ensuring compliance with reporting. The period has seen the team step up to the challenge left by staff leaving while also providing a high level of ongoing support with very little impact on the overall ANDS team.

2.8.3 Activity/Deliverables for 2012-13

Over the last reporting period due to staff changes the activity of the Project Office team has been maintaining a focus on business as usual. The core focus has been somewhat fragmented and focused on delivering a high level of internal service and support. The key activities over the last twelve months have been;

- Training new staff that have joined the team
- Providing a high level of administration and project support with a reduced staffing level
Continuing to drive a cohesive communications message to the sector using share, our e-newsletter andsUP and the social media platforms that we present

Provide training and up skilling of non-administrative staff to reduce the amount of central administration effort required.

Redeployment of tasks within project office team to ensure a high level of support

Simplify processes to be light weight to support the reduced headcount at the project office

### 2.9 Promotion

ANDS has undertaken a large number of promotional activities during the period July 2011 to the end of June 2012. These include the following.

#### 2.9.1 Presentations/attendance at Conferences

ANDS staff have presented at and/or attended a range of international and local conferences to promote the service and to establish relations with other parties. These include:

- eResearch Australasia 2012, Sydney
- Doing Data Better, Melbourne University, Melbourne
- Financial Analytics Forum, Melbourne
- The Importance of Metadata and Effective Data Management Practices for Climate Change Adaptation Research and Responses Workshop, Brisbane
- UNE eResearch Showcase, University of New England, NSW
- Launch SURE System, Perth
- USyd RDM Resources Launch, NSW
- Tools & Data in the Cloud, Canberra
- National Conference Series - Where is the Evidence? Policy, research, and the rise of grey literature, Canberra
- Simon Goudie - Wiley Data Journals, Canberra
- ARC / NHMRC / ARMS research seminar, Canberra
- Tim Berners-Lee public lecture, Canberra
- Fenner school: research data Sternstein Room, Canberra
- File management – Menzies, Canberra
- Teaching Forum Workshop-Improving student learning from research experiences in undergraduate courses, Canberra
- TERN/ANDS data management symposium, Canberra
- ANDS-NeCTAR SA Dojo, Adelaide
• ANDS-NeCTAR ACT Dojo, Canberra
• RDSI/ANDS/QCIF one day workshop, Brisbane
• Australian Coastal and Oceans Modelling and Observations Workshop (ACOMO), Canberra
• Research Impact Workshop, Australian National University, Canberra
• National Archives of Finland visit at the Australian National University, Canberra
• Victorian eResearch Strategic Initiative Showcase 2012, National Centre for Synchrotron Science, Clayton
• Science Meets Parliament, Canberra
• Research Data Alliance Meeting, Washington, USA
• EUDAT (European Data Infrastructure) 1st Conference, Barcelona, Spain
• Workshop on Adaptation Knowledge Portal Initiative, Australian National University, Canberra
• Research Data Alliance 1st Council Meeting, San Francisco
• TERN 4th Annual Symposium, Canberra
• Great Barrier Reef World Heritage Area Integrated Monitoring Framework Data Management Workshop, Townsville, QLD
• Charles Sturt University eResearch showcase, Wagga campus, NSW
• Research Data Alliance Launch, Sweden
• Climate Change Adaptation Information Hub Metadata Workshop, Brisbane
• Research Data Alliance 2nd Council Meeting, Vancouver Island, Canada
• Research Data Alliance planning meeting, Washington DC
• Australian Catholic University Research Management planning session, Sydney
• ALA Science Symposium, Canberra
• Nordic e-Infrastructure Conference, Trondheim, NO
• ISOCat Workshop, Sydney
• ReDBoX Community Day, Adelaide
• Data-driven journalism day, Melbourne
• DataCite Annual Meeting and AGM, Cologne, DE
• e-IRG Workshop, Amsterdam, NL
• SMART Dashboard Launch, Wollongong
• Nordbib Conference, Copenhagen, DK
• CSIRO Computational Scientists Forum, Canberra
• THETA Education conference, Hobart
• CSIRO CMAR, Hobart
• Data Citation Workshop, Brisbane
• Data Citation Summit, Canberra
• GeoNetwork User Forum - Office of Spatial Data Management, Canberra
• Spatial Information Modelling COP Workshop, Canberra
• Spatial@Gov Open Day, Canberra
• Australia New Zealand Metadata Working Group, Canberra

2.9.2 Forums

ANDS has hosted or presented a wide range of forums over the reporting period to build our communities, share knowledge and expertise, and provide support to our various audiences. It has changed focus from physical forums to more emphasis on virtual forums.

Virtual Events

ANDS has started hosting a variety of free virtual events, to help our partners and communities learn, discuss and exchange ideas, and meet colleagues without even leaving their desk, and they have proven to be incredibly popular. Virtual events – including webinars, virtual meetings and ‘how-to’ sessions – remove any constraints associated with location, of either the presenter or participants. This has resulted in greater diversity of presenters and topics, which is a great benefit to our local and international audiences and also allows our communities to hear from and inform international and national perspectives on a wide range of topics and issues.

Virtual events also enable members of the ANDS team to share their own expertise, with a number of staff having already facilitated sessions on topics including: Ethics, Data Management policies, Project Blogging 101 and Data Interviews. Australian institutions can now tap into, and request a virtual one-to-one session with specific staff on project support, RIF-CS and Trove support.

Increasingly we are noticing that the real value of the virtual events is how they are facilitating community building, and enabling the community to learn from each other.

Some sessions are recorded and available here: ands.org.au/presentations/audio-video.html

Data Citation Roundtables

The ANDS Data Citation Roundtable sessions (hosted virtually) have seen presenters from the Digital Curation Centre and Thomson Reuters Data Citation Index in the UK, DataCite in Germany, and Data Citation at Oak Ridge National Laboratory USA.

Researchers, research institutions and universities are rapidly becoming aware of the possible strategic value of both data citation metrics and the reuse of their data. ANDS Data Citation Roundtables are designed to extend the understanding of, and encourage discussion around the issues involved in data citation, with the ultimate aim of developing best practice guidelines and to help build a Data Citation community of
practice—especially citation using Digital Object Identifiers (DOIs). This is of interest to those wishing to ensure that their research data is correctly cited. The Terrestrial Ecosystem Research Network (TERN), Australian Antarctic Division (AAD), Australian Data Archive (ADA), Griffith University, CSIRO, Integrated Marine Observatory System (IMOS) and Geoscience Australia, have all had representatives attend the Roundtables.

Other forums ANDS participated in include:

- ANDS-NeCTAR SA Dojo
- RDSI/ANDS/QCIF one day workshop, Brisbane
- Terrestrial Ecosystem Research Network – ANDS Forum
- CSIRO Computational Simulation Scientists Workshop
- CSIRO CMAR Forum
- Australian Antarctic Division (AAD) Data Management Centre Forum

2.9.3 Consultation meetings

ANDS staff have also continued to consult extensively with potential and current partners and stakeholders to discuss the services that ANDS offers and how they might be of interest to them.

2.9.4 Newsletter

The ANDS quarterly newsletter, *share*, continues to create awareness of ANDS, its activities and successes amongst the research community and stakeholders by providing updates on ANDS-funded projects, highlighting achievements and promoting ANDS events and objectives. The themed approach to each issue that was initiated in 2011 has been continued in all subsequent issues, resulting in more focused pieces that have been widely appreciated. This has resulted in an increase in the circulation of the newsletter, for both the digital and print versions. The four issues produced over the reporting period have focused on researcher impact, research data as an institutional strategic asset, institutional transformations and researcher benefits.

2.9.5 Other activities

Twitter

In November 2011 the ANDS official Twitter account (@andsdata) was launched at eResearch Australasia. Initially launched on a trial basis, the ANDS twitter account has primarily been used to communicate with our stakeholders about ANDS events, publications and news, as well as relevant stories/information from the sector. Twitter has subsequently become a useful communication channel that complements our two main communication channels: the ANDS website and *share*.

Interaction with @andsdata from our stakeholders has grown steadily throughout the period. It is worth noting that we have a number of international stakeholders regularly interacting with @andsdata including staff members from JISC (UK), Digital Curation Centre (UK), Heather Piwowar and the University of...
Newcastle (UK). Tweets that have generated the highest number of click-throughs (meaning engagement with the content) are all promotional tweets about ANDS events, guides and share.

Figure 10: Twitter report outlining overview of growth

**Youtube**

The last twelve months has seen an increased focus on utilising our Youtube channel. It has been a great way to share webinars as well as training videos and presentations that the capabilities team have presented. Below are some statistics that demonstrate the success that we have seen with Youtube over the last twelve months.
Figure 11: Youtube report outlining overview of performance and engagement
2.10 Risk Management

ANDS maintains a Risk Register. The risk assessment methodology, adapted from the Australian Risk Management Standard AS/NZS 4360:2004, involves identifying and analysing each risk in terms of how likely it is to happen (Likelihood) and the possible impacts (Consequence). The risk score for each risk is calculated by combining Consequence score with the Likelihood score. This will give a risk score of between 2 and 10, which can then be mapped onto a Risk Scoring Matrix to give a risk rating of HIGH (8-10), SIGNIFICANT (7), MEDIUM (6) or LOW (2-5). Where there is more than one risk measurement area for scoring consequence, the highest combination of scores is taken as the final risk score.

The risk management register was reviewed in January 2013. ANDS assessed the residual risk level of all 11 risks in the register, taking into account the effect of the risk mitigation strategies that have been put in place.

At completion of this review it was concluded that there was only one change from the previous review in April 2012.

Risk 11 was re-assessed and there was an increase in the residual risk from low to high. This increased due to the funding uncertainty at the time. There was concern that this would lead to potential staff departures.

The strategy that was employed to manage this was to provide an increased level of communication to staff and regular update on funding outcomes.
3 Progress against milestones

3.1 Data Capture Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
</table>
| 12Q3           | Completion of 7 projects  
Other projects being monitored and assessed as required.  
Delivery of 70 records to Research Data Australia  
5 new data management tools available via Product catalogue | 1 additional project completed this quarter, with ongoing assessment of those projects still underway  
13,931 records added to RDA  
1 new data management tool added |
| 12Q4           | Completion of 11 projects  
Other projects being monitored and assessed as required.  
Delivery of 500 records to Research Data Australia  
10 new data management tools available via Product catalogue  
Redeployment of 2 tools | 2 additional projects completed this quarter, with ongoing assessment of those projects still underway  
662 records added to RDA  
2 new data management tools available via Product catalogues  
Redeployment of MyTardis |
| 13Q1           | Completion of 10 projects  
Other projects being monitored and assessed as required.  
Delivery of 300 records to Research Data Australia  
10 new data management tools available via Product catalogue  
Redeployment of 5 tools | No additional projects completed this quarter, with ongoing assessment of those projects still underway  
158 records added to RDA  
No new data management tools available  
No additional redeployment |
| 13Q2           | Completion of 1 projects  
Delivery of 200 records to Research Data | No additional projects completed this quarter |
<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 new data management tools available via Product catalogue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Redeployment of 5 tools</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2188 records delivered to RDA</td>
<td>No new data management tools available</td>
</tr>
<tr>
<td></td>
<td>No additional redeployment</td>
<td></td>
</tr>
</tbody>
</table>

### 3.2 Metadata Store Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>12Q3</td>
<td>All projects underway and project plans received. Projects being monitored and assessed as required.</td>
<td>All projects commenced at 22 institutions, 10 project plans received. Standard ANDS project monitoring and assessment in place for the others</td>
</tr>
<tr>
<td>12Q4</td>
<td>Completion of 2 projects Other projects being monitored and assessed as required.</td>
<td>No projects completed yet. Standard ANDS project monitoring and assessment in place for the others</td>
</tr>
<tr>
<td>13Q1</td>
<td>Completion of 10 projects Other projects being monitored and assessed as required.</td>
<td>2 projects completed. Standard ANDS project monitoring and assessment in place for the others</td>
</tr>
<tr>
<td>13Q2</td>
<td>Completion of 10 projects</td>
<td>12 more projects completed</td>
</tr>
</tbody>
</table>
## 3.3 Public Sector Data Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>12Q3</td>
<td>Engagement with agencies to deliver water data</td>
<td>Commenced work with 7 research groups to develop the Australian Urban Water Collection</td>
</tr>
<tr>
<td></td>
<td>Contract with PROV to deliver archives data from key state agencies and NAA</td>
<td>Contract with PROV initiated. Business Requirements developed and recruitment of resources by agency commenced.</td>
</tr>
<tr>
<td></td>
<td>Agreements with state eResearch agencies to deliver state public sector data</td>
<td>State Records NSW begin development of API for providing harvest to Research Data Australia.</td>
</tr>
<tr>
<td></td>
<td>Engagements with state eResearch agencies commenced</td>
<td></td>
</tr>
<tr>
<td>13Q1</td>
<td>Automated feeds of collection level data by initial 4 government agencies</td>
<td>Feeds from ALA, AAD, AIMS and e-Atlas developed, tested and implemented.</td>
</tr>
<tr>
<td></td>
<td>Engagements commenced with a further 4 agencies</td>
<td>Engagements with Geoscience Australia and Australian Ocean Data Network for improvements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Testing of PROV and State Records NSW feeds commence.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Australian Urban Water Collection (7 research groups) developed and implemented into Production.</td>
</tr>
<tr>
<td>13Q2</td>
<td>Automated feeds of data into the other discipline portals in ARDC has been provided by selected government agencies</td>
<td>Public Record Of Victoria and State Record of NSW implement production feeds. GA and AODN implement improved production feeds.</td>
</tr>
<tr>
<td></td>
<td>Automated feeds of collection level data by 4 prioritised engagements</td>
<td>Discussions held with Department of Sustainability, Environment, Water, Population and Community re working with ANDS. This department has carriage over many government agencies and is asking for ANDS assistance in managing data. Initially for its National Environment Research Program.</td>
</tr>
</tbody>
</table>
### 3.4 ARDC Core Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>2012-13 Milestones</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>12Q3</strong></td>
<td>RDA Institutional Pages and topic pages Integration of Citation Identifier Infrastructure integration – first phase</td>
<td>Contributor Pages and Topic page functionality released. 18 research organisations have created home pages. DataCite “See Also” in production</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>12Q4</strong></td>
<td>User feedback to Registration, Discovery and Data Collection Page Creation Infrastructure Vocab service phase 2</td>
<td>Change management system processed 75 user requested improvements into production Phase 2 not completed. This service was de-prioritised and release to production is now scheduled for Q1 2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>13Q1</strong></td>
<td>Integration of Citation Identifier Infrastructure second phase Research Activity infrastructure complete</td>
<td>Phase two completed: users update the URLs associated with their DOIs Partially complete: NHMRC grant records ingested; ARC ingest re-scheduled to 2013-4; grant information service partial roll out in 2013-4.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>13Q2</strong></td>
<td>ANDS Software Products re-factoring, packaging, and documentation ANDS software full integration with national services (ARC, NHMRC, GeoScience Au)</td>
<td>Complete rewrite of ANDS Registry completed. ANDS Software publicly available at github.com/au-research/ to enable collaborative development and ensure the most recent release code is readily available for use. Graphics design review of Research Data Australia website implemented. Integration complete.</td>
</tr>
</tbody>
</table>
### 3.5 ARDC Application Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>12Q3</td>
<td>25 projects started</td>
<td>25 projects started</td>
</tr>
<tr>
<td></td>
<td>2 projects complete</td>
<td>1 project complete</td>
</tr>
<tr>
<td>12Q4</td>
<td>15 projects produce early demonstrations of value</td>
<td>1 early demonstration of value</td>
</tr>
<tr>
<td>13Q1</td>
<td>13 projects complete</td>
<td>No further projects complete</td>
</tr>
<tr>
<td></td>
<td>10 remaining projects produce demonstrations of value</td>
<td>No further demonstrations of value</td>
</tr>
<tr>
<td>13Q2</td>
<td>10 remaining projects complete</td>
<td>2 further projects complete</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 further demonstrations of value</td>
</tr>
</tbody>
</table>

### 3.6 International Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>12Q3</td>
<td>ANDS appointed as DWF NGS for Australia</td>
<td>ANDS appointed (although DWF renamed to Research Data Alliance)</td>
</tr>
<tr>
<td></td>
<td>Australia's membership in DWF confirmed</td>
<td>Membership confirmed</td>
</tr>
<tr>
<td></td>
<td>Data thematic session held at the 2nd EU-AU Joint Workshop for Research Infrastructure</td>
<td>Session held</td>
</tr>
<tr>
<td></td>
<td>1-day data/e-infrastructure workshop held with Australian and European participants</td>
<td>1 day workshop held with focus on linguistic data</td>
</tr>
<tr>
<td>12Q4</td>
<td>Outcomes from 1-day data/e-infrastructure workshop funded and commenced</td>
<td>Outcomes agreed, funded and commenced</td>
</tr>
<tr>
<td></td>
<td>ODIN project funding agreed by EC and project commenced</td>
<td>ODIN project successful in funding application</td>
</tr>
<tr>
<td>Milestone Date</td>
<td>Milestone</td>
<td>Progress</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------------</td>
</tr>
</tbody>
</table>
| 13Q1           | Activity plan outlining Australia's future participation in DWF accepted by DIISRTE  
Outcomes from 1-day data/e-infrastructure workshop completed and reported  
DWF successfully commenced | Activity plan accepted  
Outcomes not reported as still ongoing |
| 13Q2           | Data contributions provided as part of 3rd EU-AU Joint Workshop for Research Infrastructure | Third workshop deferred until November |
4 Deviations from the Project Plan

There were a number of deviations to the initially submitted project plan, but none from the revised business plan as described in the letter describing the changes. These changes resulted from slowing ANDS expenditure in order to have a less dramatic reduction in services beyond June 2013, so that it was possible to maintain enough funds to run ANDS at minimum critical level for 18 months. ANDS saved expenditure through:

- salary savings as there have been some staff that have not been replaced as we move to a smaller staff, and some salary saving resulting from delayed replacement of roles that will continue
- not conducting some face to face Seeding the Commons community events
- not investing in two external contracts: to fund both the ARC and the NHMRC to provide a establish national grant information infrastructure
- not conducting some proposed commissioned reports from the Frameworks and Capabilities area
- not conducting some events face to face but instead using teleconferences and increasing our use of webinars

The consequence of these savings is that ANDS has not been able to do all that it intended. The specific consequences are:

- a reduction of deliverables with regard to the integration of the Data Citation service and the pilot Standard Vocabularies Service
- a reduction in the scale of our non-funded institutional engagement activities that we have internally labelled “better data”
- a redirected Capabilities events program
- an increase in virtual meetings such as data surgeries and webinars
- one less Applications project was initiated than was originally intended

Expenditure Reduction

Actual expenditure was $7.04 million less than budgeted; the main reasons are outlined below.

- This was partly due to lower expenditure on staff than budgeted by approximately $0.7 million:
  - Specifically a delay in appointing Partnerships and Project Managers as well as delay in appointing our Business Manager.
  - There were also some savings where staff have resigned and roles have not been replaced due to planned reduction in activity through the period.
- The bulk of the variance (approx. $4.9 million) is a result of delays in projects not reaching final milestones.
5 Commitments

In order to understand the Financial Report, we need to consider the conversion of committed (but not contracted) funds to contracted funds. This process involves getting agreement from all parties regarding the deliverables and timelines, negotiating the terms and conditions of the sub-contract and executing the contract.

<table>
<thead>
<tr>
<th>Contracts Schedule</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Contracts as at 30 June 2012</td>
<td>52,754,672.00</td>
</tr>
<tr>
<td>Total Contracts executed from 1 July 2012 to 30 June 2013</td>
<td>3,164,643.00</td>
</tr>
<tr>
<td>Monash University’s ANDS-funded projects paid via internal transfer</td>
<td>2,100,175.43</td>
</tr>
<tr>
<td>Total Contracts in progress as at 30 June 2013</td>
<td>798,500.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>58,817,990.43</strong></td>
</tr>
</tbody>
</table>

Figure 13: Progress on committed and contracted funds
## 6 Financial and Human Resources

The following table indicates ANDS expenditure by program for July 2012 to June 2013. Income is shown in the audit statement in Section 9.

### ANDS Expenditure for the Financial Year ended 30th June 2013

<table>
<thead>
<tr>
<th>Item</th>
<th>2012</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applications</td>
<td>549,895</td>
<td></td>
</tr>
<tr>
<td>Data Capture</td>
<td>390,770</td>
<td></td>
</tr>
<tr>
<td>Monash University Data Capture Project</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Metadata Stores</td>
<td>151,686</td>
<td></td>
</tr>
<tr>
<td>Monash University Metadata Stores Project</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ANDS International Infrastructure</td>
<td>64,739</td>
<td></td>
</tr>
<tr>
<td>Project Office</td>
<td>118,456</td>
<td></td>
</tr>
<tr>
<td><strong>Total Salaries</strong></td>
<td><strong>1,275,545</strong></td>
<td></td>
</tr>
<tr>
<td>Operating Expenses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Capture External Infrastructure Contracts</td>
<td>2,632,779</td>
<td></td>
</tr>
<tr>
<td>Data Capture Internal Program Expenses</td>
<td>41,038</td>
<td></td>
</tr>
<tr>
<td>Metadata Stores External Infrastructure Contracts</td>
<td>2,479,269</td>
<td></td>
</tr>
<tr>
<td>Metadata Stores External NeAT Projects</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Metadata Stores Internal Program Expenses</td>
<td>10,698</td>
<td></td>
</tr>
<tr>
<td>Public Sector Data External Infrastructure Contracts</td>
<td>1,047,739</td>
<td></td>
</tr>
<tr>
<td>Public Sector Data External NeAT Projects</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Public Sector Data Internal Program Expenses</td>
<td>3,751</td>
<td></td>
</tr>
<tr>
<td>Applications External Infrastructure Contracts</td>
<td>2,520,915</td>
<td></td>
</tr>
<tr>
<td>Applications Data External NeAT Projects</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Applications Data Internal Program Expenses</td>
<td>44,236</td>
<td></td>
</tr>
<tr>
<td>ARDC Core External Infrastructure Contracts</td>
<td>112,500</td>
<td></td>
</tr>
<tr>
<td>ARDC Core Internal Program Expenses</td>
<td>22,115</td>
<td></td>
</tr>
<tr>
<td>ARDC Core ANU Program Funding Payment</td>
<td>2,165,839</td>
<td></td>
</tr>
<tr>
<td>International Infrastructure External Infrastructure Contracts</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>International Infrastructure Internal Program Expenses</td>
<td>52,386</td>
<td></td>
</tr>
<tr>
<td>Project Office Internal Expenses</td>
<td>31,922</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total Operating Expenses</strong></td>
<td><strong>11,166,186</strong></td>
<td></td>
</tr>
<tr>
<td><strong>EIF Total Expenditure</strong></td>
<td><strong>12,440,731</strong></td>
<td></td>
</tr>
</tbody>
</table>

### Notes

1. It was noted past the close of the financial year ended 30 June 2012 that software licence fees ($975) were wrongly charged to EIF as opposed to NCRIS. This was corrected on the 10th of August 2012 through document number 401045119. This is a prior year adjustment.
Staffing and recruitment has been extensive since the establishment of the project. The organisation chart at the end of June 2013 is shown in Figure 14.

Figure 14: ANDS Organisational Chart
7 Co-Investment

7.1 Access and Pricing

All ANDS services are available free of charge to all Australian researchers at publicly funded research institutions and discovery is available to all. As such no cash has been received for the provision of ANDS services, which is in line with expectations.

7.2 Project Co-Investment

As a result of the intended timeframe for the project of 2 years, it was agreed that it would be inappropriate to require co-investment in ANDS projects. The 2009-10 and 2010-11 Business Plans identified that ANDS would place a wholly-funded ANDS staff member within an institution to achieve the aims of that institution and of ANDS. This has been changed as part of the EoI process to a more flexible arrangement where we contract with partners to allow for a mix of staffing needs. In addition, many research organisations (for example CSIRO, Monash University, Queensland University of Technology, and Griffith University) have contributed effort to the ANDS projects beyond the ANDS investment.

ANDS investments at institutions in Seeding the Commons and Data Capture activities have triggered substantial co-investment and post-investment with over $2M of institutional investment made to date, and over $3M of post-project investment. This indicates the extent to which institutions are embedding a research data infrastructure into standard operations. It is pleasing to note that this additional effort is being continued beyond the life of the projects, showing the importance being placed on research data, and the extent to which institutions are establishing coherent research data infrastructure as part of “business as usual” operations.

As a part of the agreement to fund metadata stores at institutions ANDS required that the institution indicate what co-investment it would make to demonstrate a whole-of-institution commitment to their metadata infrastructure. This amount varied across the institutions, but in total ANDS has had commitments of around $2 million in co-investment, beyond the $5 million that ANDS has provided.

To provide a perspective over the life of the project to date:

- Total ANDS Funds $35,069,000
- Total Partner co-investment $10,625,450
- Total Post project co-investment $4,118,877 (this is by necessity an estimate)

Looking just at the financial year 2012-13:

- Total ANDS Funds $7,602,628
- Total Partner co-investment $1,600,656
- Total Post project co-investment is estimated to be between 21-30% of ANDS investment
8 Performance Indicators

8.1 KPI Report

The following are the KPIs agreed in the 2010-11 Business Plan, with actual results and commentary included (some of these results have been achieved through EIF ARDC project activity, but cannot be usefully separated):

1. The number and coverage of data repositories providing metadata feeds to the national registry compared to the number of data repositories ANDS intends to build at least 80 automatic plus 100 manual metadata feeds. This will cover at least 35 out of the approximately 50 research data-holding institutions that we know about.

**Result:** 68 institutions fed collection descriptions to RDA along with 46 individual collections. From these institutions, 154 data source feeds have been setup (87 automatic and 67 manual feeds)

2. The number and coverage of institutions and number of research groups with which ANDS has engaged: ANDS will continue to engage with all Australian universities, PFRO’s, and 4 major Government data providers this year, and through them at least 50 research groups.

**Result:** ANDS is currently engaging with the following:

- 38 universities – all except the University of Notre Dame, with level based on research output
- Publicly Funded Research Organisations: ANSTO, CSIRO and AIMS, with level based on research output
- Government data providers: ANDS has engaged with over 30 government agencies apart from the PFROs. These include GeoScience Australia (GA), Australian Institute of Health & Welfare (AIHW), Australian Antarctic Division (AAD), Murray Darling Basin Authority, Bureau of Meteorology, NSW Department of Sustainability, Environment, Water, Population and Community, Queensland Dept of Employment Economic Development & Innovation, Public Records office of Victoria (PROV) and Australian Bureau of Statistics (ABS) directly; and Royal Australian Navy through the engagement with AODN; National Archives of Australia, State Records NSW and State Archives of Queensland through the engagement with PROV; and 18 museums through the Museum Metadata Exchange project including Powerhouse Museum, Australian Museum and state museums. Through the AustLII project we have exposed public data from Attorney General’s Department and various courts around Australia including the High Court. Through the project with AuScope, engagement has been with Bureau of Meteorology and various state Departments of Primary Industry and Sustainability & Environment
- National facilities: Australian Animal Health Laboratory (AAHL), Australia Telescope National Facility (ATNF), Australian Synchrotron, and research vessels: Southern Surveyor and Aurora Australis
- National capabilities: Astronomy Australia, Integrated Marine Observing System (IMOS), Terrestrial Ecosystem Research Network (TERN), AuScope, Atlas of Living Australia (ALA),
Population Health Research Network (PHRN), Australian Urban Research Infrastructure Network (AURIN), Bioplatforms Australia, Australian Phenomics Network (APN), Australian Plant Phenomics Facility (APPF), National Imaging Facility (NIF), Australian Microscopy & Microanalysis Research Facility (AMMRF), EMBL Australia, Australian National Fabrication Facility, AusBiotech, Research Infrastructure Support Services, Australian Biosecurity Intelligence Network, Australian Social Sciences Data Archive, Australian Access Federation, National Research Network, Pawsey Supercomputing Centre, National eResearch Collaboration Tools and Resources (NeCTAR), Research Data Storage Infrastructure (RDSI).

3. The number of institutions with research data management policies and practices consistent with ANDS recommendations: 25

**Result:** 12 – CSIRO, Monash University, University of Melbourne, Queensland University of Technology, Griffith University, University of Wollongong, University of Newcastle, Edith Cowan University, La Trobe University, Australian Catholic University, University of New England, and University of Canberra. ANDS has engaged with many more institutions on the development of research data management policies and practices this year, however comparatively few of those engagements have been finalized at this stage. There are eight additional institutions with draft research data management policies.

4. The number of times a search is initiated with an ANDS discovery service: There was no target for this year; this is the first year that we have taken these measurements.

**Result:** 93,796 searches (targeted search with a particular keyword) since 21 September 2011.

5. The number of times an ANDS data page is accessed: 300,000 in this year, up from the KPI of 100,000 for last year.

**Result:** 176,151 page views and 112,239 unique page views (from Google Analytics tool). Unique page views increased 64% from last year.

6. The satisfaction of researchers and partners (see below) with ANDS services as measured by an annual survey - no number can be given here, but a report will be provided.

**Result:** A Survey was completed by 83 respondents: including Partners (67), Researchers (14), Unknown (2). Over 70% of ANDS University Partners contributed responses to the Survey.

**Key Outcomes:**

| 1. Satisfaction with ANDS services overall | • 88% were satisfied or very satisfied  
• 5% dissatisfied |
| 2. Awareness of the value of data | • 90% thought data valuable or very valuable  
• 4% thought research data of little value |
| 3. Change in research data management (RDM) practices over last two years | • 57% noticed a change | 23% noticed no change |
| 4. Satisfied with Research Data Australia (RDA) | • 20% were unsure  
| 5. Research Data routinely deposited into data repositories | • 84% satisfied or very satisfied  
| | • 20% usually or always  
| | • 55% seldom |

Compared to the [last Survey](#), ANDS improved its level of satisfaction from 56% (3.93 / 7) in 2010 to 80% (3.2/4) in the 2012/13 Survey; a 43% increase during the intervening period.

7. The number of data access and sharing agreements with stakeholders – principally research institutions, government data agencies, government research agencies: ANDS aims to strike at least 30 agreements to make data available.

**Result:** ANDS has agreements with **68 organisations** to publish data collection descriptions in either the public or draft systems, up from 49 last year, 15% increase from June last year.

There are two measures that ANDS will not have full control over, but that are important and will measure our success in influencing others’ behaviour:

8. The number of research data sets in the ARDC: more than 10,000 collections

**Result:** **56,599** collections as at 30 June 2012. 34% of these collections came from Geoscience Australia (GA), 24% were from Queensland Facility for Advanced Bioinformatics (QFAB), and 27% came from Australian Ocean Data Network (AODN). The remaining 15% of the total collections are from 65 Research Data Australia contributors.

The number of ANDS research data sets increased **42%** from June last year.

9. The number of research data sets with persistent identifiers: 10,000

**Result:** **10,727** persistent identifiers (9,015 PIDS handles, 1,712 DOI handles), **56%** higher than June last year. These identifiers were minted by 17 Identify My Data (PIDS) clients and 22 Cite My Data (DOI) clients.

There is a final measure that ANDS aspires to – it will be measured but is unlikely to be a useful short-term KPI.

10. The number of times a data set is reused and referenced – the ultimate long-term measure. At present ANDS is unable to report on this. However, the DOI-based data citation service will encourage publication of data collections with persistent identifiers in a citable form, and existing citation tracking services are expanding their coverage to include such data citations.

**Notes:**

An ANDS data page is a page generated from the ANDS collections registry that describes a data set, a collection, a research group, a research project, or an institution.
ANDS will focus on monitoring Institutions that are research data producing organisations, such as the Bureau of Meteorology, Landsat, the Australian Synchrotron, the Cultural Collections sector, and the research data using organisations, such as the Universities, the PFROs, and affiliates. Many organisations have both roles.

Researchers have many partners in carrying out research and ANDS needs to satisfy their needs as well – this includes funders, assessors, institutional representatives, such as DVC-Rs, eResearch Directors, Information providers such as libraries, IT providers such as University ITS Departments, partner service providers, such as ARCS and NCI, as well as umbrella organisations such as disciplinary bodies such as the Academies, international research bodies, etc.

The qualitative measures are intended to capture not only usage figures, but also attitudinal attributes – ANDS only succeeds with cultural change, so this will be measured as well. The first survey has set benchmarks, but will also help inform future surveys.

### 8.2 Progress over the Life of the ANDS Project

Having been in progress for 42 months it is possible now to indicate life over the project, along with expectations based on current activity, and when appropriate an indication of what might be possible. Based on activity taking this year, it is possible to determine an estimate for 2012-13, and when appropriate give a figure indicating total coverage. The measures that are described augment the KPI information with additional measures that help understand the Australian Research Data Commons.

<table>
<thead>
<tr>
<th>Measure</th>
<th>09-10</th>
<th>10-11</th>
<th>11-12</th>
<th>12-13</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Institutions Participation in RDA</td>
<td>21</td>
<td>35</td>
<td>40</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Research Institutions with Data Management Policy and Practice</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>12</td>
<td>43</td>
</tr>
<tr>
<td>Institutional context capture tools</td>
<td>0</td>
<td>6</td>
<td>43</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Institutional Research Metadata Store</td>
<td>0</td>
<td>9</td>
<td>25</td>
<td>25</td>
<td>43</td>
</tr>
<tr>
<td>Research Data Provider Participation</td>
<td>NA</td>
<td>9</td>
<td>30</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Research Data Infrastructure Partners</td>
<td>4</td>
<td>23</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Data Collections</td>
<td>1,173</td>
<td>26,746</td>
<td>40,811</td>
<td>56,599</td>
<td></td>
</tr>
<tr>
<td>Research Data Exploitation Tools</td>
<td>9</td>
<td>26</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Fields of Research Coverage</td>
<td>5</td>
<td>21</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 4: Progress over the life of the project

Note 1. Research Institutions denotes all 39 Universities and 4 Publicly Funded Research Organisations
Note 2. Research Data Provider participation is measured by those organisations that are not Universities and Publicly Funded Research Organisations that have an agreement with ANDS to publish research data collections descriptions – this may be indirect through Infrastructure provider partners.

Note 3. Research Data Infrastructure partners refer to those NCRIS and EIF infrastructure providers that could exchange research data collections descriptions – this measures how effectively ANDS is partnering with other problem specific data investments.

Note 4. Research Fields of Research based on ANZSRC FOR codes – all, including DIVISION 22 PHILOSOPHY AND RELIGIOUS STUDIES are now covered in Research Data Australia (RDA).

Note 5. Research Data Australia had not been launched as at 30th June 2010.

Note 6. This denotes the total number of Research Institutions: all 39 Universities and 4 Publicly Funded Research Organisations.

Note 7. This number included organisations that we are now considering as under the auspices of a Research Institution.

8.3 Overall Progress

In previous Annual Reports we were able to report on some overarching achievements:

- The Australian Research Data Commons (ARDC) has been established, and substantial progress has been made in populating it. The ARDC is a combination of the set of shareable Australian research collections, the descriptions of those collections including the information required to support their reuse, the relationships between the various elements involved (the data, the researchers who produced it, the instruments that collected it and the institutions where they work), and the infrastructure needed to enable, populate and support the commons. In summary, all components of the ARDC exist, but not all components have been established and used at all relevant institutions, and by all relevant researchers.

- ANDS has driven a change in the research data management uptake in Australia. ANDS is engaged with all major research institutions, and importantly they are engaged with and learning from each other’s approaches.

- Research data infrastructure and research data management have been established at a significant number of research institutions. ANDS estimates that there are approximately 300 people working on data management within research institutions, which is probably a ten-fold increase compared to January 2009. Research institutions are seeing substantial value in this infrastructure (see for example quotes from the University of Adelaide and James Cook University). ANDS investments at institutions have triggered substantial co-investment and post-investment with over $2M of institutional investment made to date, and over $3M of post-project investment. This indicates the extent to which institutions are embedding a research data infrastructure into standard operations.

- Data is overwhelmingly on the agenda in research and research infrastructure, and ANDS has helped position Australia internationally. The Research Infrastructure Roadmap produced in 2011 saw data as crucial infrastructure for research, and this has been emphasised in the draft National Research Investment Plan which refers to the crucial role information and data play in enabling Australian research to tackle the key research challenges of the country. Very importantly, research institutional leaders are similarly seeing great value in the way that research data can give their researchers an advantage in research data partnerships and tackling larger research questions. This perception is mirrored internationally. There was a very strong emphasis on research data in the 2012 International Conference on Research Infrastructure in Copenhagen. Both Europe and the US have made significant investments in research infrastructure. ANDS has had an important role in ensuring that Australia has...
a leading role in international research infrastructure initiatives, particularly through the emerging Research Data Alliance, where Australia is partnering with the US and the EU. ANDS, together with some of the data-intensive capabilities, has made a significant contribution in ensuring that Australian researchers and research institutions are engaged and leading in these global trends.

- The Australian Research Data Commons has matured and grown substantially. There are twice as many collections and three times as many contributing institutions as compared to last year, and the ARDC now covers every Field of Research. Importantly, now that the Data Citation service is available researchers are availing themselves of the opportunity of publishing their research data, using minted Digital Object Identifiers being used to connect their collection description to the relevant data repository and to cite their data.

Taken together, two important changes have occurred during the year – firstly there is clearly an increased desire to publish and get credit for research data. The great interest in data citation is evidence of this, so the ability of Australian researchers to publish their data will be increasingly important. Secondly, Australia’s response to the open data agenda internationally has been made easier by having a locus of thinking about research data and having conversations, developing policy responses, and involving all relevant agencies, though ANDS policy and practice capabilities.

The major activities for ANDS as a whole in the 2012-13 calendar year were:

- Making more than 56,599 research data collection pages discoverable through Research Data Australia, Google, and other search engines harvested by ANDS at over 70 research data providing institutions.
- Substantially updating the national data services enabling better publication and discovery of research data.
- Over 30,000 archival collections from major state archives are now discoverable through Research Data Australia
- All major research institutions have installed substantial research data infrastructure, notably metadata stores, that enable them to participate in the Australian Research Data Commons.
- Demonstrating the value of reusing data with a suite of applications that has enabled high profile researchers to provide compelling demonstrations of new ways of conducting data intensive research.
- Establishing the Research Data Alliance, in partnership with the EU and the US, which is committed to “data sharing without barriers” with an initial Plenary in Gothenburg, Sweden, and attracting a very substantial level of international commitment.
9 Audit Statement

The signed copy of the following document will be delivered separately.

---

**Australian National Data Service Project - EIF Funding**
Statement of Income and Expenditure for the Financial Year Ended 30 June 2013

<table>
<thead>
<tr>
<th>Item</th>
<th>2013</th>
<th>Appendix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant</td>
<td>-</td>
<td>A</td>
</tr>
<tr>
<td>Interest</td>
<td>546,681</td>
<td>A</td>
</tr>
<tr>
<td>Other Income</td>
<td>-</td>
<td>A</td>
</tr>
<tr>
<td><strong>Total Income (a)</strong></td>
<td>546,681</td>
<td></td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salaries</td>
<td>1,275,545</td>
<td>B</td>
</tr>
<tr>
<td>Non-Salary</td>
<td>11,165,189</td>
<td>B</td>
</tr>
<tr>
<td><strong>Total Expenses (b)</strong></td>
<td>12,440,731</td>
<td></td>
</tr>
</tbody>
</table>

Surplus/(Deficit) for the reporting period (a - b) $11,894,086
Brought forward surplus/(deficit) from 30 June 2012 (c) 21,842,201
Balance Carried Forward to next Reporting Period (a - b) + (c) 9,948,115

We, Edwina Conniah and Joel Chibert, hereby confirm the following:

(i) The detailed statement of income and expenditure for the ANDS Establishment Project (shown above) represents a correct view of the financial performance for the year ended 30 June 2013.

(ii) The Funding was expended for the Project and was used in accordance with the agreement with the Department of Innovation, Industry, Science, Research and Tertiary Education.

**PROFESSOR EDMINA CORNISH**
Provost and Senior Vice President
Monash University

**MR. JOEL CHIBERT**
Director
Research & Revenue Accounting Services, Monash University

Date: \( \_\_\_\_\_\_\_\_ \)

---

30 September 2013

ands.org.au
10 Appendices

10.1 Confidential Information

There is no confidential information.

10.2 Project Description Detail

10.2.1 Data Capture Project Descriptions

<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANU – Earth Sciences</td>
<td>The Seismology group within the ANU Research School of Earth Sciences (RSES) runs an extensive monitoring and data collection program based on instruments located in various locations around Australia. The instruments log data in instrument specific formats. Data is typically recorded on SDcards and retrieved manually. The memory cards are returned to Canberra for further processing. Data is downloaded from these cards and converted to the MiniSEED format in common use within seismology. Data is then submitted to the Incorporated Research Institutions for Seismology (IRIS) repository. While the full SEED specification files, in common with many data formats, contain both data and an embedded metadata payload, MINISeed files are data only. Metadata is normally stored in dataless SEED format files; that is, SEED format files that contain the metadata payload. Data is currently stored locally on a Unix server within RSES. Collection metadata in RIF-CS format will be made available via Research Data Australia (RDA) to provide a starting point for the development of an online collection of Australian seismological data.</td>
</tr>
<tr>
<td><strong>ANU – Optical Astronomy</strong></td>
<td>The ANU’s Research School of Astronomy and Astrophysics (RSAA) operates a number facilities including Siding Spring Observatory. The RSAA has successfully installed a new instrument at Siding Spring Observatory, the wide field survey telescope (SkyMapper). To facilitate publication of SkyMapper data, appropriate sustainable infrastructure was required. Automated data publication workflows were also required to ensure the successful generation and publication of collection level metadata. This infrastructure will ensure easier access to and greater use and reuse of internationally important astronomy data sets. This project will also serve as the basis for the data publication component of a number of other internationally significant optical astronomy projects. Meteorological data from observations at Siding Springs will also be published.</td>
</tr>
<tr>
<td><strong>ANU – Phenomics</strong></td>
<td>The Australian National University (ANU) is lead agent for The Australian Phenomics Network (APN). The ANU has a number of Phenomics and Genomics research groups. Publication of this metadata to the ANU Data Commons will facilitate the reuse of the data within the ANU and the resubmission of datasets for further analyses both within Australia and overseas. The objective of this activity is to link the genomic and phenomic data sets through PODD so the metadata and data is discoverable and longevity, use and reuse of the data is ensured.</td>
</tr>
<tr>
<td><strong>ANU – Humanities and allied disciplines</strong></td>
<td>The goal of the Project was to facilitate the publication of existing data, metadata and contextual information from research work discovered during the Seeding the Commons project (SC06) and other data audits. By identifying and assisting in the publication of these datasets the Division of Information hopes to encourage the development of a culture of dataset publication. The complexities of the generation of structured datasets encoded with standard metadata may be a block to the publication of datasets within the traditionally less numerate disciplines. The Division of Information seeks therefore to build on the work of ANU Seeding the Commons project by modifying the data capture workflows to ensure the production of good quality metadata encoded according to standard schemata as an aid to data interchange and reuse. This work was also allied to the ANU’s involvement in the workspace and collections interoperability strands of Project Bamboo (<a href="http://www.projectbamboo.org">www.projectbamboo.org</a>)</td>
</tr>
<tr>
<td>CSIRO – Research Data Service Multi-Source Data Capture</td>
<td>CSIRO’s newly established Research Data Service (RDS) currently has technology that supports a small subset of domain specific data (i.e. water and pulsar data) and also generic self-serve deposit of any type of data from across the organisation. Recent requests from various business areas within CSIRO have highlighted the need to complement this capability with new functionality that enables the ongoing automatic deposit of data from various sources. In order to efficiently respond to this demand in a way that is scalable, the RDS wishes to provide this automation functionality through an enterprise focussed method that makes the addition of future ongoing data deposits a system administration &amp; configuration activity as opposed to a software development activity. The approach was to produce functionality that supports automated deposit from specified sources such as defined locations on file systems, database management systems or defined drop-box locations. The system also targets specific metadata standards for the CSIRO’s Sensor and Sensor (SSN) Network Transformational Capability Platform (TCP).</td>
</tr>
<tr>
<td>Deakin - Crystal Orientation Data Collection for Conversion to a General Data Type</td>
<td>The project was undertaken to address the fact that electron diffraction information captured by Deakin’s electron microscope facility was unmanaged, disconnected, invisible and single use. The project sought to address these deficiencies by creating an automated transformation process to enable automated data capture and to facilitate coherent metadata labelling and storage. The project promoted appropriate data management, publishing, sharing and re-use of electron diffraction information in the study of metals deformation mechanisms. The primary objective of the project was to develop infrastructure within Deakin’s electron microscope facility to support the on-going data capture and management of metals deformation data and metadata, supporting the discovery, reuse and sharing of this research data by other researchers into the future.</td>
</tr>
<tr>
<td>Deakin - Enhancing Filtration Membrane Fouling Data Collection for Water Treatment Research</td>
<td>The project was undertaken to enhance the capacities of confocal laser scanning microscopy (CLSM) for filtration membrane fouling data creation. To achieve this, the project developed sophisticated 3d image processing/analysis software. The project also generated metadata for data collections produced through the confocal laser scanning microscope and provided this data to Research Data Australia (RDA) for wider consumption.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>James Cook University – Tropical Data Hub</td>
<td>This project provides consistent format of data, long-term storage and aids re-use by providing sharing options (e.g., open access, embargoed for some time or closed access) and forcing researchers to provide appropriate metadata. The system alleviates the issue researchers have with data storage and backup. Researchers obtain long-term storage for data by simply setting up a project in the EnMaSSe system. As part of the set up, metadata about the project as well as more specific metadata about the data feeds is collected. This enables the system to not only configure the storage space and ingesters for the data but to also generate metadata records for both the project and the individual datasets that will be created. These records are then automatically loaded and published in the JCU Research Data repository and published to RDA. This allows for a much higher visibility of the datasets being collected and will hopefully lead to increased collaborations and research data re-use that in turn will lead to more citations.</td>
</tr>
<tr>
<td>Macquarie – Glycomics Repository</td>
<td>The ANDS-supported component of this initiative: ‘Linking Glycomics Repository with Mass Spectrometer Data Capture’ has seeded the infrastructure to capture, collate and disseminate the metadata on glycomics knowledge to the Australian and International research community. By leveraging the technical developments and services deployed by ANDS, over 1000 records from the GlycoSuiteDB database of glycan structures attached to proteins have been migrated to Research Data Australia. These records provide access to well-managed bibliographic references and rich metadata descriptions of glycan structures and their biological context. The data flow that connects mass spectrometry data acquisition from the Australian Proteome Analysis Facility (APAF) is now also integrating ANDS components and vocabulary services to ensure analytical data is linked to the glycan structure repository, UniCarbKB, which is now part of a new National eResearch Tools and Resources (NeCTAR) project. This initiative will allow biological and medical researchers to build upon existing efforts and will enhance research and subsequent new discoveries in glycobiology.</td>
</tr>
<tr>
<td>Monash – Comprehensive Data Management for Microscopy Research Datasets</td>
<td>The purpose of this project was to provide software tools to support the organisation of microscopy research datasets, particularly microscopy imaging data along with the associated annotation and metadata, both instrumental and experimental, into institutional repositories and into other discovery services including Australian Research Data Commons (ARDC). The solution delivered incorporates an existing platform for visualisation, management and analysis of biological microscope images – OMERO, which has been widely adopted within the international microscopy community. This platform establishes a concept for mandatory annotation of digital experimental data that will maximise the use, re-use and distribution of scientific raw data and experimental information within the research community. Additional functionality developed as a plug-in for OMERO provides the researchers with the ability to specify data license and generate files that allow microscopy data collections to be made publicly available for search and harvest by Research Data Australia. Overall, the solution provided by the Comprehensive Data Management for Microscopy Research Datasets project has enhanced the MMI microscopy imaging data management process. This in turn supports research in areas that benefit from access to these microscopy data collections.</td>
</tr>
<tr>
<td>Adelaide – Genomics Data Capture</td>
<td>This project provides a system to manage and store genomics information in a secured area. Analytic outputs of raw data are also tacked to discover re-useable analysis pipeline for a particular problem. The system has two ways to publish information for sharing. Sequence data is published to public sequence repository European Bioinformatics Institute (EBI) European Nucleotide Archive (ENA) and at the same time the collection description, along with relevant, associated party, activity and service records are made available for harvest to Research Data Australia (RDA). With this software solution, • data are managed in a model complies EBI and RDA requirements to ensure relevant information are collected; • information can be easily found; • whole analysis history of sequence which includes tools used are captured; • submission to EBI is easier with an authorised submission account with EBI and required information collected from very beginning satisfies EBI strict rules on how to describe data.</td>
</tr>
<tr>
<td>UNSW – Data Capture and Integration</td>
<td>The aim of this project was to establish an institutional research data management service that would enable the capture and curation of a diverse range of analytical data from instruments within UNSW. This project has not focused on a single research discipline or field but has been a collaboration between the Mark Wainright Analytical Centre, the Schools of Chemistry, Chemical Engineering and Medical Science, all of whom have contributed instruments now connected to the new data management service, named ACData. The Mark Wainright Analytical Centre comprises major instrumental facilities used by researchers from Science, Engineering and Medicine faculties at UNSW. The facilities include instruments, such as microscopy, NMR, X-ray fluorescence and diffraction and proteomics mass spectrometry. Due to the diverse range of users who access the centre and their instruments, the Analytical Centre has been the hub of the project, including the NMR and spectroscopy instruments housed there. In addition to these instruments, Potentiostat instruments within the School of Chemistry, a Capillary Porometer from the School of Chemical Engineering and Slide Scanners from within the School of Medical Sciences have also been integrated within ACData.</td>
</tr>
<tr>
<td>Newcastle – Data Capture for the Data Commons</td>
<td>This project produced a web-based system by which researchers can define and conduct surveys, and then easily publish the existence of the survey dataset in Research Data Australia (RDA). The system is entitled QuON, and at present is largely being used by researchers in the area of health sciences. QuON supports a range of survey question types, and supports definition of complex branching conditions so that participants’ navigation of the survey can be made dependent on previous responses. The kinds of researcher-participant interaction is intentionally generic, so QuON is not aimed at any particular research discipline and is suitable for use in conducting any survey.</td>
</tr>
</tbody>
</table>
| University of Queensland – Spatially Integrated Social Science | The SISS project focussed on developing online tools that allow researchers to quickly access rich Australian socio-spatial datasets related to voting outcomes and census data, conduct statistical modelling, and visualize spatial relationships in the data. The project:  
- Established a repository of statistical variables derived from Australian Bureau of Statistics Census Data and Australian Electoral Commission voting data;  
- Developed geospatial visualisation and statistical services for analysing these variables;  
- Exposed these visualisation and analysis services through three Web portals that allow researchers to easily analyse the variables; and  
- Shared RIF-CS research data collection descriptions of the derived data via ANDS Research Data Australia (RDA). |
| University of Queensland – Microscopy/Microanalysis Image and Data Repository | The Centre for Microscopy and Microanalysis (CMM) bridges the sciences. It seeks to understand the structure and composition of all materials at atomic, molecular, cellular and macromolecular scales. CMM is a foundation member and the Queensland Node of the Australian Microscopy and Microanalysis Research Facility (AMMRF). The primary goal of the MIRAGE project was to assist researcher’s using CMM facilities in storing and managing their microscopy data. The project team implemented systems that could:  
• Capture instrument metadata, and any available administrative metadata  
• Transfer data from instruments into the MIRAGE repository  
• Allow users to browse and reorganize their datasets in the repository  
• Allow users to view images or upload the files and process them  
• Allow users to upload files produced by analysis of their raw data. The project’s secondary goals were to assist and encourage the sharing and reuse of research data captured using CMM’s facilities and to add a level of accountability and integrity to data collected from the CMM. |
| University of Queensland – DIMER Diffraction Image Repository | This project has implemented and deployed an automatic data capture component within the UQ ROCX facility that allows diffraction image datasets to be harvested directly from each instrument. The owner of the dataset receives an email notification and simply needs to complete a small number of metadata fields in order to create a complete record. Files are directly transferred into DIMER, removing the need for users to manually upload large diffraction image datasets. The automatic capture of diffraction image datasets from the UQ ROCX facility improves the integration of DIMER into the researcher’s workflow. This is vital given the time constraints faced by researchers: scientists acknowledge the value of depositing datasets into open-access repositories, in terms of enhancing the credibility of published findings by allowing experiments to be reviewed and potentially repeated; however, given the pressure on scientists to progress their laboratory work and, in particular, to publish their work in academic journals, the publication of datasets is often overlooked. In addition to making datasets available to the wider community following publication, storage in DIMER also enables researchers to fulfil their internal data management obligations. DIMER provides a repository that is accessible to all X-ray crystallography researchers at UQ, with support for granting read and/or write access to fellow laboratory group members, supervisors, and other collaborators. This facilitates collaborative research and prevents datasets becoming “lost” when a researcher leaves UQ. In the absence of using DIMER, researchers may store images on portable disks or their local machine. By automatically capturing datasets in DIMER, it is ensured that datasets are reliably stored and backed up. |
### University of Queensland - 3D Anthropological and Archeological Data capture of 3D digital models

The project successfully converted the UQ Anthropology Museum’s existing offline catalogue into an online catalogue containing around 30,000 items accessible via a Web-front end: [http://catalogue.anthropologymuseum.uq.edu.au/](http://catalogue.anthropologymuseum.uq.edu.au/). This Web interface provides public access to:

- Metadata descriptions of (most) items in the museum’s collection. Some item descriptions are not publically accessible for copyright or cultural reasons.
- Digital photographs of selected items. At the time of writing this report there were 1693 items with digital photographs in the catalogue, but this will grow as the museum’s digitisation efforts continue. Most of these items have multiple digital photographs depicting the item from different angles.
- 3D scans of selected items. There are currently 13 objects with associated 3D high resolution scans.
- Digital representations of documents associated with the collection such as donor correspondence, condition reports and collector biographies.

### University of Sydney – Metadata Store Aggregator

An institution-wide metadata store will be implemented in collaboration with University of Sydney ICT. This store will aggregate information from suitable University enterprise systems, have applicability across many areas of data capture and complement the Sydney Seeding the Commons project.

### University of Sydney – ExCite9

The ExSite9 tool has been designed to support the description of highly structured or unstructured (multimedia) data captured out in the field. The tool has been designed to enable researchers from any domain to describe their data at the collection level (in support of stronger high level description for the purposes of discovery and archiving for the purpose of access and reuse). Researchers that collect and load data from still or moving image cameras, sensors, etc in the field can range in domain from: linguistics and anthropology through to geology, botany and archaeology. Their shared and critical need is to collate and curate data on a desktop, apply the appropriate and domain specific metadata (to support their research) and to transfer that to external hard drive or to a data repository in an efficient, standardised and secure manner, for further processing and potentially publishing for citation and reuse.
| **University of Tasmania - Redmap** | Redmap is a citizen science project and associated web site in Tasmania for collection of data on species distribution extensions related to climate change. This project extends the Redmap Concept nationally. The project will enable significant rebuilding of the web site and database to allow for increased complexity collecting data across a large geographical area, increased complexity in data processing (including data validation) and automated metadata and data delivery to the TPAC MEST and ANDS Collection Registry. The resulting system:  
  a) Allows sightings of selected marine species around Australia to be entered by members of the public.  
  b) Allows expected species distribution to be managed in the form of disjoint polygons.  
  c) Automatically compares sightings against expected distributions.  
  d) Automatically delivers sightings metadata to the TPAC MEST.  
  e) Expansion of the existing website to encompass collection of data on a national level  
  f) Allows capture of additional species-specific data where impacts due to range extension/variation are considerable (eg. recording the habitat surrounding the sighting, as for Centrostephanus rodgersii).  
  g) Integrates with existing data visualization tools.  
  h) Allows direct download of all datasets through redmap.org.au by any visitor (with provision for embargoes or restrictions on selected materials).  
  i) Provide a semi-automated workflow for incorporation of data validation and quality control mechanisms. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UTS – Maximising the benefits from data-intensive processes at UTS</strong></td>
<td>The aim of this project is to develop and put in place systems to facilitate the management, sharing and re-use of data from data-intensive processes at UTS. The project is comprised of three activities. The first will provide metadata capture at data acquisition time from the Microbial Imaging Facility's microscopes. The second will extend the Labshare system software to support remote access to equipment, acquired data, and the recording of metadata, by researchers. The third will provide access to processed Mars imagery data and associated metadata for researchers in the Planetary Sciences.</td>
</tr>
</tbody>
</table>

### 10.2.2 Metadata Stores Project Descriptions

See body of document
10.2.3 Public Sector Data Project Descriptions
See body of document

10.2.4 Applications Project Descriptions
See body of document
## 10.3 Progress against activities

### 10.3.1 Data Capture Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANDS has either entered into contracts (or has substantially agreed on project descriptions) for Data Capture projects at 24 institutions. As noted above, some projects have been, or will be completed before the planning period. Shaded projects are still being defined at time of writing, but are expected to be agreed by the beginning of the planning period.</td>
<td>At June 30, 2012, ANDS had either entered into contracts for all of the Data Capture projects at all of the EOI institutions. This totalled 70 projects, of which 33 are now complete. A breakdown of the progress made in relation to this is provided in section 2.2.2, with projects completed in the current reporting period described in section 2.2.3</td>
</tr>
<tr>
<td>The remaining funds in the Data Capture program will be allocated according to two criteria:</td>
<td>Excess funding was redirected to other ANDS programs, in consultation with the Steering Committee.</td>
</tr>
<tr>
<td>Data capture infrastructure that is required to deliver data relevant to identified themes. This will require the Steering Committee to agree on the themes and accept recommendations from the Executive Director that particular infrastructure investments are required to deliver critical data.</td>
<td></td>
</tr>
<tr>
<td>Data capture infrastructure that is required to meet the needs of one of ANDS’ demonstrations of value or research champions. This will require the Steering Committee to agree on the instruments or researcher(s) selected and accept recommendations from the Executive Director that these particular infrastructure investments are required to support the success of this approach.</td>
<td></td>
</tr>
</tbody>
</table>
### 10.3.2 Research Metadata Store Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>All of the agreed 2010-11 activities that were originally funded through the Early Activity component of the EIF ARDC Project Plan are expected to complete.</td>
<td>All these projects have been completed, with the exception of the Griffith Research Metadata Hub Upgrade, which is still to deliver one section of documentation.</td>
</tr>
<tr>
<td>ANDS anticipates that a number of institutions will be reluctant to deploy two new pieces of software to manage collection-level and object-level information. Institutions may also prefer to have an integrated view of their research data outputs. ANDS therefore intends to fund the development of an expanded and generalised Combination Store, suitable for wider deployment, as well as selected deployment in a small number of reference institutions.</td>
<td>ANDS funded the further development of the ReDBox software through QCIF. Uptake of this software has been stimulated through the funding of metadata stores projects. 14 institutions have indicated they will use ReDBox. Other institutions have also indicated they will use VIVO (5), ORCA (1) or MyTARDIS (1), all of which have received some ANDS funding.</td>
</tr>
<tr>
<td>For the Object Store solution, ANDS staff will first analyse existing solutions (both ANDS-developed and available from elsewhere) to identify the best fit for the required functionality. ANDS will then consider the most cost-effective and sustainable way to deliver the total functionality required. If any adaptation or extension is required, ANDS intends to approach development partners who have demonstrated their readiness, willingness and ability in past engagements rather than undertake an open call. This is because of the need to have this solution in place soon.</td>
<td>ANDS opted not to commission development in this area in order to release funds for institutional collection metadata stores.</td>
</tr>
<tr>
<td>Installation of new pipes will be undertaken using a selection mechanism that starts with the EOI round ranking, but is also informed by the number of possible deployment candidates for any solution. Preference will be given to those solutions that will meet the needs of the largest number of the most research-intensive organisations. The same criteria for development partners will be used as for the adaptation/extension of the Object Store solution.</td>
<td>ANDS opted not to commission development this area in order to release funds for institutional collection metadata stores. Pipes are being developed within institutions as part of these projects.</td>
</tr>
</tbody>
</table>
Proposed Activity

Deployment support activity will be undertaken using a mix of ANDS-funded staff and local e-Research services providers. For institutions, ANDS will use the existing ranked list developed for the EOI round and progressively work down (assuming institutions are ready, willing and able to engage). ANDS staff will make this selection informed by the amount of funding available, the preparedness of institutions to adopt the solutions, the extent to which any deployment will support the four ANDS Transformations, and the estimated deployment costs. In addition, ANDS will engage with RDSI nodes (once selected) to assist them to provide both Object and Collection level management of metadata, and associated feeds to the ANDS Collections Registry. This will be done in alignment with the results of the RDSI selection process.

<table>
<thead>
<tr>
<th>Agency or Institution or Project</th>
<th>Status and Description</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>AuScope SISS Deployment</td>
<td>Deployment of the Spatial Information Services Stack (SISS) into 10 government and state or territory agency data providers including CSIRO, Geoscience Australia, BoM, ODSM and DPI(Vic) and to release 73 spatial information data collections into the Australian Research Data Commons. A relationship has been established with VeRSI as a Victorian node to provide support to Victorian deployments of SISS and work is underway on the identification and capture of datasets from the Bureau of Meteorology.</td>
<td>Auscope completed all contracted obligations and in addition, implemented an instance of SISS at the National Computational Infrastructure (NCI). The super compute capability of the NCI will enable researchers to access, process and manipulate large data sets provided via SISS. The available data sets include Geoscience Australia’s back catalogue of landsat imagery. Also developed by this project was a Virtual Exploration Geophysics Laboratory (VEGL), which is a workflow engine, enabling researchers to process geophysical data. The workflow utilises the Amazon Cloud. During the project, SISS was implemented into 12 Government agencies, with a total of 103</td>
</tr>
<tr>
<td>Date</td>
<td>Section</td>
<td>Text</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>30 September 2013</td>
<td>CSIRO Water Resources Observation Network</td>
<td>5 significant collections of data from the Murray Darling Basin Sustainable Yields projects have been exposed via Research Data Australia. Eight open source technology tools associated with the project were released for re-use, including: one to enable the ingest of data and metadata into the CSIRO Data Management System, a RIF-CS harvesting tool, a software library that provides programmatic access to the ANDS Persistent Identifier (PID) Service, NetCDF conversion tools, establishment of data storage and serving capability and a data and metadata management infrastructure. Additionally a web interface to provide data consumer access to data held within CSIRO was established and will be implemented for access to other data from the organisation. Work from this and other ANDS projects has been leveraged into the establishment of an enterprise wide data management and delivery system and service.</td>
</tr>
<tr>
<td>30 September 2013</td>
<td>AODN Data from National Research Vessels</td>
<td>This project has delivered 900 core data sets captured from research vessels (Southern Surveyor and Aurora Australia) published via the Australian Ocean Data Network portal and Research Data Australia (RDA) with further feeds of other data from the AODN portal under review. These data collections will expose the data of 6 commonwealth agencies with primary responsibility for marine data. Subsequent to the completion of the project a further 9000 datasets have been captured. The Australian Ocean Data Network (AODN) and ANDS have developed a strong partnership. AODN has provided over 12000 marine science data collection descriptions to the ANDS Registry. AODN is an aggregator of data collected by the Australian Marine Community. Primary datasets are contributed by the Integrated Marine Observing System (IMOS) and the six Commonwealth Agencies with responsibilities in the Australian marine jurisdiction (Australian Antarctic Division,</td>
</tr>
<tr>
<td>Australian Legal Information Institute (AustLII)</td>
<td>The original commitment has been met with over 400 databases and collections from courts tribunals and government agencies exposed via RDA. These data sets are principally the ‘raw materials’ of most legal research: legislation of all forms; Court and Tribunal decisions; treaties; official materials interpreting legislation; and reports proposing law reform. Exposure to AustLII databases provides the only free access, comprehensive national view of legal data of this nature. Additionally exposure has been given to LawCite, a completely automatically-generated case citation service. An extension has been granted to December 2011 – to enable the gathering of a further 50 significant data collections and feeds of this data are underway. The AustLII team are also keen to network further, with for example, NCJRDN, to provide an even more comprehensive view of Australia’s legal data. They have also obtained a LIEF grant to digitise backsets. This data will also be contributed to Research Data Australia and the ARDC and when</td>
<td>Completed</td>
</tr>
<tr>
<td><strong>Museums Metadata Exchange</strong></td>
<td>This project has coordinated metadata from 18 museums across Australia including both major museums such as Powerhouse, state museums and the Australian Museum as well as smaller regional museums. The project will deliver a metadata exchange which will gather metadata from these museums and automate a feed into RDA. A test ingest of initial records into RDA test area to facilitate exposure of aggregated museum and history collections has already taken place. It was anticipated that on completion 700 collection records would be published. It delivered over 1000 collections. As well as the metadata store collecting metadata from other museums and the tools to automate the feed to RDA, the project will establish a vocabulary tool to support data standardisation in the museum sector.</td>
<td>Completed.</td>
</tr>
<tr>
<td><strong>National Archives of Australia</strong></td>
<td>Proposal reviewed and value of commitment revised. NAA are re-assessing their intentions.</td>
<td>See PROV below</td>
</tr>
<tr>
<td><strong>Public Records Office of Victoria</strong></td>
<td>A proposal is being developed to identify collections a feed metadata from the state based archives agencies as well as the National Archives of Australia. This is currently under negotiation. If successful it will provide a comprehensive view of Australian archives data.</td>
<td>The Public Record Office of Victoria (PROV) completed a contracted project during this period. PROV lead the project on behalf of the Council of Australasian Archives and Records Authorities (CAARA). Government archives across Australia hold massive quantities of documents which are of great value to researchers, particularly in the humanities and social science fields. The project developed software tools for the automated extraction of metadata about these holdings and their transformation into the format required for delivery to the ANDS Registry.</td>
</tr>
</tbody>
</table>
| GeoSciences Australia | This engagement has enabled the automated exposure of data holdings from the GeoMet and GeoCat data catalogues and has resulted in an initial release of approximately 700 data collections which was soon followed by a feed bringing it to 7300 collections. It has enabled the identification of relevant data to be fed to other discipline portals, for example marine data. Automation of feeds has leveraged off the SISS deployment at GA. Work continues on the development of ANZSRC classification to allow greater manipulation of the data collections. | PROV have provided access to over 14000 Collections to date. The software tools and other products developed are open source and reusable by other National and State archives and agencies. State Records NSW have implemented components and have provided over 15000 Collections to the ANDS registry. Queensland State Archive have expressed interest in implementing the solution. There is potential for an Archives National Collection. Targeted archives are National Archives, State Records of South Australia and State Records of Western Australia, with whom relationships have been established. Metadata from GeoMet and GeoCat catalogues (via Geonetwork metadata catalogue) has been harvested into the ANDS registry. 19342 Collection records and 3330 related Party records. The Collection records have been enriched to include ANZSRC classification codes (mapped from the GA subject keywords).

GA are in the process of constructing request XML for the minting of Digital Object Identifiers from DataCite, via ANDS’ services. The Collection records are currently being further enriched to contain high-quality citation metadata so that DataSets can be accurately cited. Record citation metadata will be sent to Thomson Reuters Web Of Knowledge to be indexed within the Data Citation Index. |
### 10.3.4 Australian Research Data Commons Core Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Data Collections Registration Infrastructure:</strong>&lt;br&gt;Multi-schema harvest and transformation support (xml storage)&lt;br&gt;Downloadable software package and support for independent database layer.&lt;br&gt;Integration with Location Infrastructure and example code snippets for partners&lt;br&gt;Integration with FOR vocabulary and example code snippets for partners&lt;br&gt;“Community source” development support</td>
<td>Multi-format support implemented.&lt;br&gt;Auto crosswalk engine implemented for NHMRC. Additional ability to store native xml.&lt;br&gt;Database layer extracted with rewrite of ANDS registry.&lt;br&gt;Packaged location widget client to access the Gazetteer service.&lt;br&gt;FOR vocabulary widget created and made publicly available with demonstration and release guide.&lt;br&gt;ANDS Software publicly available at github.com/au-research/ to enable collaborative development and ensure the most recent release code is readily available for use.</td>
</tr>
</tbody>
</table>

| **National Data Collection Discovery Infrastructure:**<br>‘See Also’ linking: DataCite, NLA, ALA.<br>Enhanced subject browse: “Field of Research”<br>Integration with Location Infrastructure<br>Downloadable software with independent indexing layer.<br>“Community source” development support | See Also linking to DataCite implemented; NLA and ALA de-prioritised and on hold.<br>FOR vocabulary widget integrated to allow users to browse the vocab tree.<br>Location widget integrated into infrastructure.<br>ANDS Software publicly available at github.com/au-research/ to enable collaborative development and ensure the most recent release code is readily available for use. |

<p>| <strong>National Data Collection Description Infrastructure:</strong>&lt;br&gt;Topic pages&lt;br&gt;Institutional Party pages | Topic page generation implemented and topic pages now included in search results within RDA.&lt;br&gt;Institutional party pages implemented as “Contributor” pages allowing institutions to create, maintain and... |</p>
<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visualise the data commons (contributors map; connections mapping)</td>
<td>publish a party records.</td>
</tr>
<tr>
<td>Graphic design update</td>
<td>Visualisation de-prioritised and on hold</td>
</tr>
<tr>
<td>Integration with Location Infrastructure</td>
<td>Modern and user-friendly Research Data Australia website implemented.</td>
</tr>
<tr>
<td>Downloadable package</td>
<td>Location widget integrated into infrastructure.</td>
</tr>
<tr>
<td>“Community source” development support</td>
<td>ANDS Software publicly available at github.com/au-research/ to enable collaborative development and ensure the most recent release code is readily available for use.</td>
</tr>
</tbody>
</table>

**Dataset Identifier Infrastructure:**

ANDS Cite My Data service integrated with existing ANDS services: allocation of DOIs to already registered collections and simultaneous DOI allocation and collection registration.

In collaboration with the International Infrastructure program, continue to contribute to the design and implementation of global data identifier and citation system

Analysis of simultaneous registration of DOIs and collections conducted – implementation deprioritised and on hold. ANDS Cite My Data service enhanced to include a graphical interface to a listing of a client’s DOIs, their activity log and to view their registered application settings. Interface also allows user to update the URL of their DOIs.

Continued participation and contribution to DataCite, ORCID, and ODIN

**Place Names Infrastructure:**

ANDS is in partnership with responsible government agencies (GA/OSDM) to establish a robust national infrastructure that will allow place names to be validated by both individuals and software systems against an Australian Gazetteer (a directory that lists names of geographical place and features and includes spatial co-ordinates)

Phase 2 (marine, boundary boxes, expanded dataset)

Phase three planning (historical names, vernacular names, crowd sourcing...)

Development and implementation of example code which queries The Gazetteer of Australia 2.0 to provide spatial location resolution for place names.

The code was implemented in the Register My Data screens in the Registry to assist users in entering spatial coverage.

Phases two and three de-prioritised and on hold.
<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrate Gazetteer services with existing ANDS services</td>
<td></td>
</tr>
<tr>
<td><strong>Researcher Identification Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>NLA party identifier project benefit realisation stage</td>
<td>NLA party identifier retrieval function transported to redesigned Registry system. Party records continue to be created from SRU calls to the NLA party registry. NLA SRU call-backs implemented on ANDS demonstration system to perform call backs from the NLA test system.</td>
</tr>
<tr>
<td>Trial / adopt new directions: ORCID.</td>
<td>ORCID researcher lookup widget created and integrated into existing ANDS services.</td>
</tr>
<tr>
<td>Integrate Researcher Identification Infrastructure with existing ANDS services</td>
<td></td>
</tr>
<tr>
<td>In collaboration with the Australian Research Council and the National Health and Medical Research Council create software components and services that will improve the accessibility and quality of information about research activities undertaken within Australia. When completed, the project will enable the public to discover and track the research undertaken through ARC and NH&amp;MRC funding and to follow its connections to research outputs, both nationally and internationally.</td>
<td>Latest NHMRC activity records created in the Data Commons. ARC re-scheduled to 2013-4 Connections between ARC and NHMRC activity records with collections commenced. Implemented multi-schema harvesting into the ANDS Registry allowing funding information to be regularly harvested from the ARC and NH&amp;MRC in their native publication format (CSV).</td>
</tr>
<tr>
<td><strong>Standard Vocabularies Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>Build on the existing prototype to enable the creation, management, and publication of human and machine readable ‘terminologies’ (also known as controlled vocabularies) for use by the Australian innovations sector. Maintain an external advisory group (technical working group) to receive input from government</td>
<td>Australian and New Zealand Standard Research Classification (ANZSRC) Fields and Research (FOR), Australian and New Zealand Standard Research Classification (ANZSRC) Socio-economic Objective (SEO) and Registry Interchange Format – Collections and Services (RIF-CS) vocabularies published</td>
</tr>
<tr>
<td>Proposed Activity</td>
<td>Progress</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>and research stakeholders</td>
<td>for use by the Australian innovations sector.</td>
</tr>
<tr>
<td><strong>ARDC Infrastructure Establishment:</strong></td>
<td>Specialist technical support provided through ANDS service desk, specialist DOI infrastructure implementation support, specialist Party Infrastructure implementation support, and participation in webinars, events and developer “dojos”;</td>
</tr>
<tr>
<td>Provide technical assistance and design support to those organisations building the distributed infrastructure of the ARDC</td>
<td>Maintenance of ANDS service desk and associated systems and processes</td>
</tr>
<tr>
<td>Provide specialist technical backup to members of the Australian research and higher education sector in the uptake of ANDS infrastructure</td>
<td>Participation in the ANDS international engagement program</td>
</tr>
<tr>
<td>Continue the establishment of ANDS infrastructure to support those services which have been implemented to date</td>
<td>Adoption support provided to TERN and TDH.</td>
</tr>
<tr>
<td>In collaboration with the International Infrastructure program, continue to work to enable integration of ARDC Core Infrastructure with international infrastructure networks (such as the National Science Foundation and JISC)</td>
<td></td>
</tr>
<tr>
<td>Facilitate adoption of ANDS software solutions</td>
<td></td>
</tr>
</tbody>
</table>

### 10.3.5 Australian Research Data Commons Applications Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tropical Data Hub and Tools Development</strong></td>
<td>The Tropical Data Hub (TDH) is a platform to serve data sets related to Tropical research from a single virtual location and will include data-integration, visualisation and analytical tools to enable researchers, managers and decision-makers to collaborate around the data. The project has created a set of complementary tools consisting of server side data manipulations and web accessible interfaces that are designed to process</td>
</tr>
</tbody>
</table>

30 September 2013  
ands.org.au
<table>
<thead>
<tr>
<th>Project Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species Distribution Records</td>
<td>This project developed a tool that reuses data available through the ALA and the JCU Tropical Data Hub to explore the potential impacts of climate change on a wide range of bird species in Australia; and engage in improving our understanding of the species and the modelling of species distributions. There is currently a scarcity of transparent online tools which integrate species distribution data, locality data with climate change scenarios in an integrated fashion which will facilitate the modelling of current and future species distributions based on climate scenarios. This tool does this for bird species.</td>
</tr>
<tr>
<td>Climate Change Impacts Downscaling tool</td>
<td>Now renamed as <strong>Adaptations and Impacts Downscaling tool</strong>, this project built tools that identify suitable downscaled climate data sets and read the contents of those data sets. A dialog with four key impacts research groups defined their data needs for their research, and developed the tools to generate these data, as required by the impacts researchers. The data can be exported in formats defined by the impacts researchers to enable their research.</td>
</tr>
<tr>
<td>Marine Video</td>
<td>This project has supported the development of a database framework applicable to marine imagery that will facilitate the timely delivery of standardized, quantitative estimates of ecologically relevant indicators (such as absence/presence, percent cover, abundance and distribution of benthic</td>
</tr>
<tr>
<td>Project</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Health Impacts of Climate Change</td>
<td>Now renamed as <strong>Extreme Weather and Health Impacts</strong>, this project developed software that enhances existing research data infrastructure with tools that merge population, health and environmental data for analysis and inference in environmental epidemiology. It built a system to publish specific workflows online that integrate datasets and produce an analysis assessing and projecting the health impacts of extreme weather events.</td>
</tr>
<tr>
<td>Climate Change Adaptation Information Hub</td>
<td>This project has built a software system (terranova.org) that acts as a central information hub for researchers in the Climate Change Adaptation research domain. The system allows users to deposit research data with associated metadata descriptions into a central managed storage infrastructure. The system also enables a variety of search types (including spatial and temporal searching) of metadata in addition to other discovery tools to locate relevant research data stored in the information hub.</td>
</tr>
<tr>
<td>Impacts on Ports Infrastructure</td>
<td>Now renamed as <strong>Decision support toolkit for climate resilient seaports</strong>, this project has extended existing applied research activity funded by NCCARF (Enhancing the resilience of seaports to a changing climate), and added value through the development of an online decision support toolkit. This can be readily used by port operators.</td>
</tr>
<tr>
<td><strong>BPA X-omics tool</strong></td>
<td>Now renamed as <strong>Validation of genomes and transcriptomes with proteomic data</strong>, this project has enabled comparison of predicted genome function against the full protein readout of the genome – i.e. the ‘proteome’. This has been difficult to do until now as few if any tools co-analyse these data in any comprehensive and time-efficient way, and the tool that has been developed in this project will greatly facilitate these analyses.</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Data-Models-Papers interconnections</strong></td>
<td>This project, demonstrated at eResearch 2012, showed the feasibility of connecting data, publications and software models using DOIs to provide the linking.</td>
</tr>
<tr>
<td><strong>Satellites to Soils</strong></td>
<td>Now renamed as <strong>Soils to Satellites</strong>, the project has developed a tool to integrate plant and animal species occurrence/community data; plant and soil genetics data; and vegetation cover data (sourced from The Atlas of Living Australia, TERN/TREND, AusCover and Bioplatforms Australia) from sites situated along an environmental gradient, moving from cooler/wetter to warmer/drier. This environmental gradient can be used as a proxy for predicted climate change. Furthermore, phylogenetic analysis of these data can indicate the relatedness of the communities and turn over in community composition along environmental gradients. The tool was demonstrated at the ALA Science Symposium earlier this year.</td>
</tr>
<tr>
<td>Project Name</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mouse Brain Map</td>
<td>Now renamed as <strong>Brain Mapping National Resource</strong>, this project has built a distributed web based system that allows the inspection, annotation and analysis of multi-modal imaging data (MRI, classical histology) at multiple scales from multiple sources. This has been applied initially as a tool for neuroscience research to investigate the mouse brain, where a major challenge associated with the combined viewing of this data is that it simply is not possible to display the full resolution of all the data modalities in any one plane with current display hardware.</td>
</tr>
<tr>
<td>Multiscale Kidney Imaging</td>
<td>Now renamed as <strong>Multimodal kidney imaging data integration</strong>, this project has developed a tool to identify and quantify the functional units of the kidney (i.e. the “glomeruli”) in three different imaging modalities: classical histology (the current gold standard method which can only be performed in autopsy); MRI imaging and CT imaging (both of which can potentially be performed on living subjects).</td>
</tr>
<tr>
<td>Biology NeCTAR Project</td>
<td>Now renamed as <strong>Cancer Genomics Linkage Project</strong>, this project has developed software that will enable researchers and clinicians to access, integrate and transform unique and valuable cancer genomics datasets from the International Cancer Genome Consortium (ICGC) and other sources and to analyse them using the power of the closely aligned NeCTAR-sponsored Genomics Virtual Laboratory software system.</td>
</tr>
<tr>
<td>Project Name</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Human Chromosome 7 proteomics</td>
<td>Now renamed as <strong>Human Proteome Browser</strong>, the tool that has been developed in this project integrates a wealth of currently publicly accessible, yet disparate protein data and functions as a portal to the human proteome for the global research communities to greatly expedite the Human Proteome Organisation (HUPO) international effort. The tool is being evaluated as one of the candidates for international adoption by HUPO.</td>
</tr>
<tr>
<td>Public Open Space modelling</td>
<td>Now renamed as <strong>POSITIVE PLACES: Spatial Analysis of Public Open Space</strong>, this project has integrated heterogeneous data sets from POS data layers with other geo-spatial build environment information and demographic and socio-economic status data, and developed a web based geospatial data analytic and visualisation tool that can support research and urban planning practice.</td>
</tr>
<tr>
<td>SMART Infrastructure dashboard</td>
<td>Now renamed as <strong>SMART’s Multi-Utility Dashboard – Infrastructure Analytics for Integrative Research</strong>, this project has developed a ‘multi-utility dashboard’ that offers infrastructure analytics based on data provided by public agencies and private operators. The online dashboard acts as a ‘one-stop-shop’ portal accessing, formatting, analysing and making publicly available information on water, energy, waste, communication and transport distribution or management networks in a given area. This information, crossed with relevant figures from demographics and economics constitutes a robust foundation for</td>
</tr>
<tr>
<td>Project</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Founders &amp; Survivors</td>
<td>Now renamed as <strong>Founders and Survivors: Genealogical Connections</strong>, this project has developed open-source software infrastructure based on the database application to enable historians to record and analyse the genealogy of an entire particular population group of interest as asserted from birth, death, marriage and other events. Importantly, the evidentiary sources for all asserted events are also recorded in the database. The project has extended the application's capability to record, in addition to the usual relationships which constitute family history, various attributes of interest to researchers, and to extract cohorts, events and data of interest for further analysis in integrated and external visualisation, life course narrative, and traditional statistical analysis tools.</td>
</tr>
<tr>
<td>CODCD/BMRI demonstrator</td>
<td>Now renamed as <strong>Brain and Mind Research Institute data integrator</strong>, this project has developed a tool to create and manage persistent links relating together research data, contextual data and clinical data (de-identified in the case of human subjects). Integration of these managed data types enables new types of statistical analyses to be run across the combined research scans and clinical records (such as the case of comparing a research brain scan and the results of a clinical cognitive test).</td>
</tr>
<tr>
<td>Marine NeCTAR Project</td>
<td>Now renamed as the <strong>Marine Virtual Laboratory Information System</strong>, this project has built on and enhanced the software infrastructure developed by</td>
</tr>
<tr>
<td>Project</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>TERN/ACEAS Project</td>
<td>Now renamed as <strong>A Data Transformation and Model Calibration System for Carbon and Nitrogen Dynamics in Australian Ecosystems</strong>, this project has produced the Semaphore tool for the (Semi-) automatic analysis of Australian ecosystem dynamics. Predictive modeling of ecosystem dynamics can be extremely labour-intensive. Raw data needs preparation to be used as input in modeling software. Predicted results are then painstakingly compared to historical data to calibrate the software to local conditions. Semaphore is an attempt to transform this manual activity into a semi- or fully-automated process. Automation reduces the possibility of human error when processing and analysing the data. It also helps provide provenance information to trace and validate predictions and model calibrations.</td>
</tr>
<tr>
<td>Urban Climate Modelling</td>
<td>Now renamed as <strong>AURIN &amp; ANDS – North West Metropolitan Region of Melbourne Data Access, Integration and Interrogation and Demonstrator</strong></td>
</tr>
<tr>
<td>Projects</td>
<td>This project has facilitated access to a myriad of data sets for the Melbourne North West corridor via the AURIN portal. The project has demonstrated the benefit of providing open access across government datasets to researchers, planners and policy makers in dealing with problems of space, place, and liveability. The value of the project has been demonstrated through four projects which have covered built environment and health, housing affordability, economic productivity, and transport and sustainability.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Catchment to Coast</td>
<td>Now renamed as <strong>SEQUITOR: A Demonstration Integrated Coastal Knowledge Platform</strong>, this project integrates observational (e.g. monitoring) data from catchment, river and coast with models of catchment, estuarine and coastal processes to enable the discovery and generation of new knowledge and better understand the catchment-to-coast system. The resulting infrastructure is now being extended by TERN to model the Logan river floods.</td>
</tr>
<tr>
<td>TERN/eMAST Project</td>
<td>Now renamed as <strong>Primary production in space and time</strong>, this project has fused disconnected data sources—weather, remotely sensed land-surface observations, CO₂ and water flux measurements, hydrograph data and remotely sensed CO₂ concentrations—to generate a time-varying field of gross primary production (GPP, the most fundamental function of all ecosystems) across the Australian continent. The resulting software infrastructure allows...</td>
</tr>
</tbody>
</table>
different ecosystem models to be compared with one another and with data, and a specific realization—a near real-time GPP hindcast at 0.05° resolution—with proven, simple light use efficiency and water balance models at its core.

10.3.6 International Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establishing the Research Data Alliance</td>
<td>ANDS was appointed the Australian organisation that helped establish RD-A</td>
</tr>
<tr>
<td></td>
<td>RD-A initial meetings conducted in Copenhagen and Washington to establish support for the RD-A approach</td>
</tr>
<tr>
<td></td>
<td>RD-A was successfully launched in March 2013</td>
</tr>
<tr>
<td></td>
<td>A number of RD-A working groups and interest groups have been established</td>
</tr>
<tr>
<td></td>
<td>The RD-A Council, establishment group and Secretariat all have ANDS participation</td>
</tr>
<tr>
<td>EU-AU Research Infrastructure meeting conducted with a data over-arching theme.</td>
<td>Event successfully held in July 2012 in Brussels with substantial follow-up leading to meeting at ICRI where RD-A is agreed upon.</td>
</tr>
<tr>
<td>ODIN – an EU project involving ANDS connecting data and parties through ORCID and DataCite</td>
<td>ANDS contributed all agreed milestones on time.</td>
</tr>
<tr>
<td>Maintaining international cooperation</td>
<td>Meetings conducted with partners in the US, UK, Netherlands, Finland, Norway, New Zealand.</td>
</tr>
</tbody>
</table>
10.4 Risk Register

ANDS maintains a Risk Register. The risk assessment methodology, adapted from the Australian Risk Management Standard AS/NZS 4360:2004, involves identifying and analysing each risk in terms of how likely it is to happen (Likelihood) and the possible impacts (Consequence).

The key risks for ANDS in executing the Projects and the risk management strategies to be employed can be grouped into four major categories.

10.4.1 Political and Governance

Risk 1 – That there are persistent negative perceptions of the Project among funding agencies and influential groups leading to a lack of buy-in

Risk Factors:

- A particular project does not have the confidence of a subsection of a community.
- Lack of confidence in governance, management, or Project delivery.
- Perceptions of slow engagement with areas of the sector.
- Change of emphasis with regard to the policies around publicly funded research data.
- Lack of certainty of the funding of the function of ANDS.
- International engagement is halted as a result of limited support of ANDS.

Risk Mitigations:

- The communications plans have been updated to ensure that the specific research communities have input into specific projects and their outcomes before, during and after the projects are undertaken.
- Diagnostic strategies have been implemented to mitigate against failure.
- Use a central point where progress of the ARDC is being tracked by metrics such as number of collections available, and numbers of datasets accessed, and the status of every project is tracked.
- Clearly articulate the Project’s message and brand.
- Engage actively with communities to avoid perception (or reality) of not meeting its needs.
- Ensure that the Project reflects the Government’s expectations through constant dialogue.
- Maintain close contact with key DIISCCRTE officers to ensure they provide input to decision making, including having an observer on the Steering Committee.
- ANDS communicates the message about the longer term vision of the function of ANDS in the sector.
- Working with funding agencies on future plans for investment in the function of ANDS.

Risk 2 – That the ANDS Project is not managed effectively

Risk Factors:
Lack of effective mechanisms for planning, leadership and management.
The structure of ANDS has a negative impact on coordinated delivery of required activities.
Collaboration between the Project and across locations is not effective.
EIF funding guidelines do not allow for sufficient Project staff to administer funded programs of work.
State based staff have competing priorities and insufficient oversight.
Projects have insufficient time to complete.
Managers departing.

Risk Mitigations:
Management and planning processes have been put in place that include formal reporting and regular reviews to ensure the efficient conduct of the Project.
Regular meetings of Project staff are held to build a team approach. Communication structures in place to facilitate working together.
Staffing levels are monitored and adjusted as required.
Contracts and partnerships with state based organisations that host Project staff have been put in place that ensure that staff are clear about their role. Ensure that ANDS-funded staff based in organisations who are ANDS sub-contractors are not placed in a position of conflict of interest.
Ensure timely projects commencement.
Ensure all late starting projects are closely managed.

Risk 3 – That the continued emphasis on external contracted engagements represents too big a burden on the lead agent

Risk Factors:
University processes, focused on student and supplier engagement, are not a good fit for sector wide activities. ANDS’ role as a sector wide agent in many of its programs has imposed additional requirements on the lead agent causing pressure on its staff to assist ANDS.
ANDS EOI approach generates clusters of work with tight timelines that impact on specific university functions such as the Solicitors’ Office and Finance.

Risk Mitigations:
Approval has been obtained for streamlined approaches at Monash University to enable ANDS to work more effectively.
Fund additional staff or specific work at Monash University to enable ANDS to work more effectively.
ANDS reduces the number of projects that are externally funded
10.4.2 Relationships

Risk 4 – That the Project's external stakeholders are not effectively engaged

Risk Factors:

- Stakeholders are not prepared to undertake the changes within their own organisations that are necessary for the realisation of the ARDC.
- Stakeholders do not see their interests in data management and those of the Project as being aligned.

Risk Mitigations:

- Maximise the effectiveness of connections between the Project and related eResearch and other initiatives, including involvement of groups outside ANDS in the ANDS Policy Forum, the ANDS Technical Forum, and the ANDS Content Forum.
- Ensure that ANDS’ engagement with stakeholders meet their research data ambitions as well as ANDS’ requirements.
- Ensure ongoing, strong engagement with the Research Sector, including research infrastructure capabilities.
- All activity plans were developed after consultation with relevant stakeholders.
- Membership of the Steering Committee includes key stakeholders.
- Performance measurement for the Project should include effective stakeholder engagement.
- Effective communication of benefits to stakeholders.
- Provide a clear rationale behind the decision process for project funding.
- Communications activities have been increased to create awareness of the value of ANDS’ activities.
- ANDS effort has been increased in creating partnerships as compared to contracting.

Risk 5 – That the Project's partners do not appropriately contribute to the Project

Risk Factors:

- Partner produces outcomes of low quality or does not meet the requirements of the contract.
- Partner expends funds in a way that is not consistent with the EIF guidelines.
- Lack of effective arrangements in place to ensure the contracted services are provided to an agreed service level.
- Service providers see themselves as disconnected from the Project’s decision-making or strategic planning.

Risk Mitigations:

- Collaboration Agreement is in place to manage output and management of joint venture partners.
Formal procurement processes have been implemented to ensure that the requirements are understood and that potential suppliers meet the set criteria.

Provide ongoing contract management to ensure the delivery of required outcomes to the contracted service levels.

Effective vendor and partner engagement approaches have been put in place.

**Risk 6 – That ANDS is not perceived as a long-term partner and hence the services are not taken up**

**Risk Factors:**

The impending end of ANDS NCRIS and EIF funding together with the different purposes of CRIS and NCRIS 2 funding, causes a perception that ANDS initiated services will not continue.

**Risk Mitigations:**

- ANDS gained approval to expend existing funding over longer timelines (consistent with other Super Science funded activities).
- ANDS creates reliable sustainable services that are offered over the longer term by other long term service providers.
- Securing CRIS funding and mapping ANDS services through this new funding regime will preserve long-term services.
- Strong contribution to DIISCCRTE Roadmap, RDIC and NRIP processes will be a mitigating factor.

**Risk 7 – That there is confusion about role of ANDS versus other related service providers in the e-Research sector which impedes effective service delivery**

**Risk Factors:**

- ANDS and eResearch infrastructure partners’ offerings are confused by possible users.
- Relationship between ANDS and state-based eResearch providers (such as Intersect) is not clear to users.
- Greater expectation of collaboration between eResearch infrastructure partners based on Research Data Infrastructure Committee (RDIC) report.

**Risk Mitigations:**

- Ensure that ANDS’ communications to a range of stakeholders provide greater clarity about ANDS services.
- Ensure that ANDS’ offerings are clearly targeted and that this is clearly stated.
- Seek greater clarity from other eResearch service providers about their offerings, avoiding either actual or perceived overlap with ANDS’ offerings.
- Increased coordination of offerings by eResearch service providers through eResearch Infrastructure.
- Discussion with NCI, NeCTAR and RDSI taking place to ensure clarity of eResearch service offerings.
Ensure RDIC provides guidance for improved communication.

10.4.3 Impact

Risk 8 – That data providers/federators do not make their data available

Risk Factors:

- The storage needs of researchers are not met, so will not consider sharing their data.
- Researchers do not wish to share their research data.
- Confidentiality agreements prevent researchers from making their data available.
- Existing data federations see insufficient value in making their data available.

Risk Mitigations:

- Strategically promote incentives and rewards in the research system for data publishers
- ANDS will co-ordinate with RDSI and Institutional stores to mitigate this risk.
- Enable data citation so that researchers get recognised for the publication of their research data.
- Encourage the use of access controlled data stores.
- Ensure that ethics agreements balance confidentiality with openness.
- Recommend that funding be linked to the provision of data via the ARDC as it becomes available.
- Provide targeted assistance to data federations to assist with integration into the ARDC.

Risk 9 – That re-users of research data do not use ANDS Services to discover, access and exploit data

Risk Factors:

- The various strategies for exposing data in the ARDC do not result in the data being easily discoverable.
- Access control mechanisms are too restrictive or complex.
- Other sources of data for re-use are more attractive or easier to use.

Risk Mitigations:

- Ensure a nuanced and multi-faceted approach to exposing the Project’s accessible data.
- Work with AusGOAL and the Australian Access Federation to identify a simple set of licensing and standard access control policies.
- Ensure that it is easy to re-purpose ARDC accessible data.
Risk 10 – That the standards and technologies that ANDS adopts are not adopted more widely

Risk Factors:

ANDS is the only user and maintainer of actual or de facto standards, leading to inability to share maintenance and development costs.

ANDS is the only source of development activity on particular technologies (RIF-CS, ORCA, ANDS Handle code).

Risk Mitigations:

Promote community ownership of standards such as RIF-CS, for example through community-led advisory boards.

Seek international engagements and partnerships to take up standards and technologies favoured by ANDS and share development load.

Ensure enough people are trained on the standards and technologies that ANDS is adopting to support wide adoption.

Make implementation decisions such that ANDS is not dependent on particular standards and technologies, but on general approaches that can be transferred across technologies.

Encourage the use of ANDS-developed technologies by other data aggregators such as Terrestrial Ecosystem Research Network (TERN).

10.4.4 Resourcing

Risk 11 – That high quality staff are hard to recruit and retain

Risk Factors:

Limited availability of skilled staff (both within ANDS and in ANDS-funded projects) impacts ability to perform tasks funded by ANDS.

Funding uncertainty leads to potential for staff departures.

Risk Mitigations:

Build a vision for the function of the ANDS for the longer term and communicate this to staff.

Provide as much certainty to staff as is possible, and involve them in navigating the future.