## Contents

### Australian National Data Service (ANDS)

1. **PROJECT STATUS**
   - 1.1 Background: 3
   - 1.2 Major Activities, Breakthroughs, Highlights, and Issues: 5

2. **ACTIVITIES UNDERTAKEN**
   - 2.1 Research Infrastructure: 7
   - 2.2 Data Capture: 8
   - 2.3 Metadata Stores: 16
   - 2.4 Public Sector Data: 19
   - 2.5 ARDC Core Infrastructure: 22
   - 2.6 ARDC Applications: 31
   - 2.7 International Infrastructure: 38
   - 2.8 Project Office: 40
   - 2.9 Promotion: 42
   - 2.10 Risk Management: 45

3. **PROGRESS AGAINST MILESTONES**
   - 3.1 Data Capture Infrastructure: 46
   - 3.2 Metadata Store Infrastructure: 47
   - 3.3 Public Sector Data Infrastructure: 48
   - 3.4 ARDC Core Infrastructure: 49
   - 3.5 ARDC Application Infrastructure: 50
   - 3.6 International Infrastructure: 50

4. **DEVIATIONS FROM THE PROJECT PLAN**

5. **FINANCIAL AND HUMAN RESOURCES**

6. **CO-INVESTMENT**
   - 6.1 Access and Pricing: 54
   - 6.2 Project Co-Investment: 54

7. **PERFORMANCE INDICATORS**
   - 7.1 KPI Report: 55
   - 7.2 Progress over the Life of the ANDS Project: 58
   - 7.3 Overall Progress: 59

8. **AUDIT STATEMENT**

9. **SUMMARY OF THE CONDUCT OF THE PROJECT**

10. **EXPECTED FUTURE USAGE OF THE INFRASTRUCTURE**

11. **EXPECTED FUTURE TRENDS**

12. **APPENDICES**
   - 12.1 Confidential Information: 65
   - 12.2 Progress against activities: 65
   - 12.3 Risk Register: 71
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 thus 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 had a high impact on the activities of ANDS, and the second change substantially affected the ANDS project as agreed in the 2012-13 Business Plan. As each of these changes occurred, ANDS continued to manage the ANDS and ARDC projects together, as they are strongly co-dependent. This report describes activity taking place against the agreed 2013-14 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 RDSS 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 a changing focus within ANDS. The advent of the RDSS initiative meant that there were new opportunities 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].

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

A consequence of the CRIS investment was that ANDS undertook 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 brought 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.

This report is the final annual report of the EIF ARDC project. It stands alone, with an initial summary of the activity of both NCRIS and EIF projects in the final year in the context of the overall project investment, before detailed description of program activities.

Figure 1: Relationship between Programs
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, nor 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 2013-14 business year saw a gradual transition from the completion of the NCRIS and EIF projects to undertaking CRIS and NCRIS 2013 activities. The activities under NCRIS and EIF were principally to:

- Maintain and strengthen existing ANDS national data services.
- Continue institutional research data management engagement.
- Continue activity to strengthen the research data policy environment.
- Continue to increase the number of collections discoverable in Research Data Australia,
- Strengthen the National Collections available to researchers.
- Conclude the program of institutional research data infrastructure establishment – including automated metadata capture, metadata stores, and tools to support data management.
- Conclude the demonstration of how data can be used effectively using the Australian Research Data Commons.
• Developing the Research Data Alliance – in partnership with the EU and the US – which is committed to "data sharing without barriers", particularly by holding the Plenary in Dublin, Ireland, and attracting an increasing level of international commitment.

In this period ANDS has transitioned from 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.

Two activities worthy of particular note during this year are:

• A comprehensive response to the Australian Research Council (ARC) changes in funding guidelines with regard to research data – ANDS provided individual advice and help, generated guides, held workshops and hosted meetings that enabled the ARC to respond directly to institutional queries.

• A deepening of the ability for researchers to publish their research data, with persistent identifiers for the data, and rich connections to publications, research grants, and researchers.

Thus, as previously noted, by the completion of the NCRIS and EIF projects, ANDS has worked across the whole sector in partnership with major research organisations and NCRIS facilities. Significant progress has been made in enabling improved data management, connectivity, discoverability, and usability by:

• Establishing the Australian Research Data Commons, a network of shared data resources.

• Populating the Australian Research Data Commons with 100,000 research data collections.

• Dramatically improving institutional research data management capacity.

• Helping to establish institutional research data infrastructure.

• Co-leading the establishment of the Research Data Alliance, improving international data exchange.

This has meant that Australian researchers, research institutions and the nation are at the forefront of the opportunities inherent in global research data intensive activity.
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 12.2. 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.

As well a program focused on **International Infrastructure** was introduced previously 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 2013-14 financial year.
2.2 Data Capture

2.2.1 Overview of program

The Data Capture program aimed 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 achieved this aim by augmenting and adapting existing data creation and capturing 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 was to partner with leading research groups and Super Science initiatives to augment or adapt data creation and capture systems.

The resulting infrastructure components included 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 also adopted/adapted/developed 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. In 2013-2014 the last 19 Data Capture projects were completed.

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.

We love citizen science projects and we’re especially keen on the Redmap (Range Extension Database Map)! This is a portal and a community where you can share your sightings of fish and other marine life as they are spotted in new habitats around Australia. The project began here in Tasmania and has since expanded around the country.

EDITOR, TASMANIAN GEOGRAPHIC, 1 AUGUST 2014
<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Title</th>
<th>Project agreed</th>
<th>Contracted</th>
<th>Underway</th>
<th>Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian National University</td>
<td>Earth Sciences</td>
<td>X</td>
<td></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></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>ANSTO</td>
<td>Metadata management for neutron beam instrument data - a joint project with Australian Synchrotron</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Australian Synchrotron</td>
<td>Metadata Capture and Storage for the Three Mature Beamlines at the Australian Synchrotron - a joint project with ANSTO</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CSIRO</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>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Curtin University of Technology</td>
<td>Curtin deployment and configuration of Institutional Metadata Repository and Research Data Portal</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Deakin University</td>
<td>Filtration Membrane Fouling Data Collection for Water Treatment Research</td>
<td>X</td>
<td></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>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Flinders University</td>
<td>Automated measurement of the responses of wildlife populations to climate change</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Griffith University</td>
<td>Smart Water</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Adult Stem Cell &amp; Neurobiological Microscopy Instrumentation and Research Data Management</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>James Cook University</td>
<td>Tropical Data Hub</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>La Trobe University</td>
<td>CMSS RLI Metadata Capture and Publication</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Glycomics Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Papyrion Data Capture</td>
<td>X</td>
<td>X</td>
<td>X</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></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Biomedical Data Platform (Molecular Biology)</td>
<td>X</td>
<td></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></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></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></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Data Publication to Interferome (MIMR/Interferome)</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Comprehensive Data Management for Microscopy Research Datasets</td>
<td>X</td>
<td></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></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Biodiversity</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>B150 Big Jam</td>
<td>X</td>
<td></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></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Adelaide</td>
<td>Genomics Data Capture</td>
<td>X</td>
<td></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></td>
<td>X</td>
<td>X</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>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 Sharing Procedures and Long-term Reuse</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</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 Archaeological 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 South Australia</td>
<td>Development and testing of a data capture tool for instruments at the Ian Wark Research Institute</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Sydney</td>
<td>SHED – Sydney Harbour Estuary Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>NSW TARDIS Node [Decided to change focus to ….]</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>X</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>X</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 UTAS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Technology, Sydney</td>
<td>UTS Data Capture</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</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>X</td>
</tr>
<tr>
<td></td>
<td>Archaeological Rock Art Data Capture</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Marine Ecology Video Capture and Storage</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Climate Change and Energy Research Facilities (CCERF)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Western Sydney</td>
<td>Climate Change and Energy Research Facilities (CCERF)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Wollongong</td>
<td>Remote Sensing Spectral Library</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 1: Status of Data Capture projects

* = the project TARDIS NSW delivered the requested software, implemented this and brought this up to production standard. However in the closing stages of the project a new solution became available through the Synchrotron that offered a better solution and in a more sustainable way. After careful consideration it was mutually decided not to complete the project.
A detailed description of the projects can be found below.

<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANU – Earth Sciences</strong></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 SD cards 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 barrier 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><strong>CSIRO – Research Data Service Multi-Source Data Capture</strong></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><strong>Deakin – Crystal Orientation Data Collection for Conversion to a General Data Type</strong></td>
<td>The project was undertaken y 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><strong>Deakin – Enhancing Filtration Membrane Fouling Data Collection for Water Treatment Research</strong></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>
</tbody>
</table>
### Project Outcomes

<table>
<thead>
<tr>
<th>Project</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>James Cook University – Tropical Data Hub</strong></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><strong>Macquarie – Glycomics Repository</strong></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 this 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><strong>Monash – Comprehensive Data Management for Microscopy Research Datasets</strong></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>
</tbody>
</table>
| **Adelaide – Genomics Data Capture** | This project provides a system to manage and store genomics information in a secured area. Analytic outputs of raw data are also tackled 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  
| **UNSW – Data Capture and Integration** | 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. |
The pulsar archive from the Australia Telescope National Facility is one of the highest cited datasets in Australia according to Data Citation Index. Researchers across the globe have cited the dataset, including the US, Canada, Europe and China.

<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
<tbody>
<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  
- 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 |
<p>| 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 &quot;lost&quot; 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. |</p>
<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
</table>
| University of Queensland – 3D Anthropological and Archaeological 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 publicly 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 (e.g. 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. |
| UTS – Maximising the benefits from data-intensive processes at UTS | 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. |

TERN and the ALA would like to thank ANDS for having the vision to support this project. The project team is very proud of what was achieved and we think it demonstrates what can be done when data is brought together.

PETER DOHERTY, ATLAS OF LIVING AUSTRALIA PROJECT MANAGER, SOILS TO SATELLITES
2.2.3 Activity/Deliverables for 2013-14

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/

- James Cook University – Tropical Biomechanics Data Capture Project.
- Macquarie University – Papyri Data Capture.
- University of New South Wales – Australia and New Zealand Neonatal Network (ANZNN).
- University of Newcastle – Data Capture for the Data Commons.
- University of South Australia – Development and testing of a data capture tool for instruments at the Ian Wark Research Institute.
- University of Sydney – Sydney Harbour Observatory Data Capture.
- University of Sydney – Agriculture and Environment Data Manager.
- University of Sydney – Microscopy Data Capture and Deposit.
- University of Sydney – Metadata Store/Aggregator.
- University of Tasmania – Redmap.
- University of Technology Sydney – UTS data capture.
- University of Western Australia – Metadata Repository.
- University of Western Australia – Rock art data studies data management.
- University of Western Australia - Organisation of Australian Underwater Video and Still Imagery.
- University of Western Australia - Integrated data capture for characterization and analysis.
- University of Western Sydney – Climate Change and Energy Research Facilities (CCERF).
- University of Wollongong – Biomechanics Data Capture Project System.

Please note that some of these projects have been reported on in the report for last year, however they had not yet delivered their final report at that point. These final reports came in this year.

2.2.4 Program Highlights, Issues and Breakthroughs

The main highlight of the previous year has been the finalising of the remaining 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 in their final phase. The challenge was to find the right level of engagement of researchers to undertake the final activities in the project. The project often had already delivered the toolset for the researcher to support them in their research. All they needed to do was undertake the last steps to get the project signed off. ANDS has identified these issues and put in place improved project management. As the Data Capture systems have only been established through this program, the right balance can be tricky. 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’s requirements and the complications that come from software development. ANDS has worked closely with partners to try and keep projects on track, and has succeeded in overcoming the issues and delivering the projects.

Out of all of these projects only one project was not finalised. The project TARDIS NSW delivered the requested software, implemented this and brought this up to production standard. However in the closing stages of the project an alternative solution became available through the Synchrotron that offered a better solution and in a more sustainable way. After careful consideration it was mutually agreed not to complete the project.

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 allowed the consequent delays in successful project completion – this cost more in ANDS project engagement, but sometimes the longer period led to better outcomes.

Learning: Projects teams may sometimes benefit from 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 better supported to enable them to make steady progress, without having them feel like they are being-overregulated. Finding the right balance can be tricky.

As the Data Capture systems have only been established through this program, it is important to keep the systems running, capturing data and passing the information on. ANDS works to keep up the engagement with the institutions after project completion to encourage them to keep the systems active and connected. An important aspect in this regard is whether there are institutional policies and guidelines in place at the organisation, for example whether Data Management Plans are required. This institutional framework has been built in parallel with the Seeding the Commons and Metadata Stores projects.
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 file-orientated, 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 are also pipes between metadata stores and data stores, and between metadata stores and the ARDC Core infrastructure.

So, software that was developed or deployed by the Metadata Stores program needed 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 needed 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 appeared, in part because of the ways in which ANDS funded activity at institutions, involving four kinds of stores:

- **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 made available through RDSI, or an international disciplined-focussed data store (such as the PDB or EMBL/EBI).
2.3.2 Outline of projects

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.

By June 2013 out of the 23 Data Capture projects ten were already completed. There were still projects running at:

- Deakin University.
- Flinders University.
- LaTrobe University.
- RMIT University.
- Swinburne University of Technology.
- University of Melbourne.
- University of New South Wales.
- University of Newcastle.
- University of South Australia.
- University of Technology Sydney.
- University of Western Sydney.
- University of Tasmania.
- University of Western Australia.

2.3.3 Activity/Deliverables for 2013-14

The focus of the work in 2013 and 2014 was in assisting the universities in completing their projects.

2.3.4 Program Highlights, Issues and Breakthroughs

A 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.

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.

Deakin University and Flinders University worked together and pooled funding to extend Redbox functionality to build a data management planning tool at the front end where the data sets are registered. This made the registering an integral part of the whole process of research data management. Another valuable outcome was creating a community of interest of Steering Committee members. People who are members of these Committees are people of influence and power. As Project Teams wound up, these people have stayed on and are still part of ANDS’ primary constituency.

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 became 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.

An issue that all of the Metadata Stores projects have had to deal with has been the integration with the Trove NLA Identity

Ands.org.au
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.

At the University of Technology, Sydney

in the Metadata Stores project they did
not only set up a metadata store but also
brought together the feeds from three
data capture projects into the Metadata
store so these are now stored centrally in
the University before being passed on to
Research Data Australia.

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).

Universities are now regarding research
data as much more important and they see
the need to have a tool to register and plan
their research data management properly.

Universities are now building processes and
policies in their institutions building on the
work started in the Seeding the Commons
(see NCRIS report) and Data Capture and
Metadata Stores projects and frequently
using the connections and steering
committees set up in these projects.

---

**Figure 2:** The dependencies between the different types of metadata stores and other systems.
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 provided the infrastructure necessary to ensure that feeds of data collection descriptions are made available from a range of public sector agencies. Identified agencies included 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, were also in scope. ANDS also needed 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 explored ways to incorporate public data collected by citizens, through exemplar projects.

The key deliverable from this program was 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 varied across agencies according to their existing infrastructure and the types of data being made available. In all cases there was 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.

Subsequently the Public Sector Data Program, and the NCRIS funded National Collections Program were integrated and jointly described in the 2013-14 Business Plan as the National Collections Program.

This report provides a report on the National Collections Program, partly funded by the EIF and partly by the NCRIS funds. The allocation of funds is described in the financial section of the report.

2.4.2 Outline of projects

In determining these national collections, this program worked with all parts of the research and public sector, including:

- Engaging with selected data custodian on establishing rich national collections.
- Working around the initial themes of Urban Water, Climate Change Adaptation, Great Barrier Reef, TERN and ARC/NH&MRC grants, the collections team will engage with research and public sector institutions to identify data collections for publication within the context of the topic and the tools associated with them.
- In consultation with research and data providers, develop an inventory of collections both institutionally and nationally significant.
- Identifying contributions to nationally significant distributed collections themes Liaise with RDSI nodes and data providers to establish appropriate storage and access.
- Engage with service providers to provide access to tools associated with data collections.
- Richly describing national collections and publish in RDA and other relevant portals.
- Enhancing RDA to provide on the fly presentation of tools and services for national collections and click through access to those services.
- Working with NeCTAR and RDSI to ensure that their online services also utilise data discovery in RDA.
2.4.3 Activity/Deliverables for 2013-14

The following table describes the current status of the projects in this area.

<table>
<thead>
<tr>
<th>Agency or Institution or Project</th>
<th>Description and status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban Water</td>
<td>A flagship distributed collection in Research Data Australia. See: <a href="http://researchdata.ands.org.au/australian-urban-water-collection/267201">http://researchdata.ands.org.au/australian-urban-water-collection/267201</a> Engagement was facilitated through the Urban Water Research &amp; Development Coalition and the Partnership Working Group. Work has continued on developing the Australian Urban Water Collection in RDA with a number of meetings held with potential contributors to the collection. A revised approach to our engagement strategy was developed based on leveraging existing relationships with ANDS Partners (data providers).</td>
</tr>
<tr>
<td>TERN</td>
<td>Work has continued with TERN to structure their metadata records so as to represent data publishing from each TERN Facility and bring this together to provide an overall TERN view in Research Data Australia. Assistance has been provided with describing entities and relationships, establishing metadata feeds, metadata management and metadata crosswalks.</td>
</tr>
<tr>
<td>Geoscience Australia</td>
<td>Work has continued with Geoscience Australia (GA) to capture citation metadata for over 20K collection records. Building on this activity was assisting GA to implement a Digital Object Identifier minting service, currently in Production. GA also recently released a new metadata catalogue requiring ANDS to regression test the integration between the two technologies.</td>
</tr>
<tr>
<td>National Computational Infrastructure</td>
<td>There has been significant activity with the National Computational Infrastructure including providing assistance with developing a data management policy, a metadata schema and the implementation of a Digital Object Identifier minting service.</td>
</tr>
<tr>
<td>Bureau of Meteorology</td>
<td>Meetings were initiated with the Bureau of Meteorology to identify additional collections for the Australian Urban Water Collection in RDA. While a number of potential collections were identified, there would be substantial additional work for the Bureau to provide these as a metadata feed to RDA. Advice was provided to the Bureau on best practice for metadata management going forward.</td>
</tr>
<tr>
<td>Population Health Research Network</td>
<td>Informal agreement was reached with PHRN for harvesting of metadata to RDA.</td>
</tr>
<tr>
<td>Commonwealth Department of the Environment</td>
<td>Input was provided to the Department’s Data and Information Management Policy for the National Environmental Research Hub. The Department will use the policy to ensure researchers comply with data management guidelines. ANDS will continue to work with the Department to provide metadata creation, data management and data repository advice to its funded activities.</td>
</tr>
<tr>
<td>Public Records Office of Victoria and State Archive NSW</td>
<td>Public Records Office of Victoria and State Archive NSW were supported in their production and implementation of an ANDS-funded solution, resulting in over 30K Collection records. This unique collection spans from the mid 1830s to today, and includes a significant amount of digitised data. Discussions were held with QLD State Archives, who would like to implement the re-usable solution, enabling the QLD archival data to be discoverable via Research Data Australia.</td>
</tr>
<tr>
<td>NSW Office of Environment and Heritage</td>
<td>ANDS was approached by NSW Office of Environment and Heritage to assist with the development of a data management policy and procedures for their agency. ANDS facilitated meetings between them and other clients to share data management journeys and associated solutions. This will be an ongoing engagement.</td>
</tr>
<tr>
<td>Australian Institute of Marine Science</td>
<td>ANDS and AIMS worked together on several fronts to re-establish the metadata feed from AIMS to RDA. ANDS also worked on the AIMS style sheet to capture citation metadata. When AIMS implements the style sheet, citation metadata will be ingested and displayed on RDA.</td>
</tr>
<tr>
<td>e-Atlas</td>
<td>The National Collections program has been working closely with e-Atlas to improve the quality of their metadata within the ANDS Registry; most notably, this has included the addition of citation metadata and improved representation of relationships between research objects. This work has been re-used to make similar improvements to the metadata from AIMS and IMOS.</td>
</tr>
<tr>
<td>IMOS</td>
<td>Through the Australian Ocean Data Network (AODN), IMOS have added an additional 2000 metadata records to RDA.</td>
</tr>
<tr>
<td>Data.gov.au portals</td>
<td>ANDS undertook analysis and technical activity to enable harvesting of metadata from CKAN. The CKAN technology is used by data.gov.au to host metadata and data. CKAN is also used by other government portals. Work started with data.gov.au When the data.gov.au ingest has been proven, work with the other portals will commence. Informal agreement was reached with NSW Dept. of Finance and Services (oversees <a href="http://data.nsw.gov.au/data/">http://data.nsw.gov.au/data/</a>) for harvesting of metadata to RDA.</td>
</tr>
<tr>
<td>Thematic linkages to collections in RDA</td>
<td>Following the release of technical and software infrastructure to support the development of theme pages, 8 new pages representing thematic views of content, were released to RDA in June. The themes align with the NCRIS 2011 Strategic Roadmap for Research Infrastructure and include: Integrated Biological Diversity, Astronomy, Terrestrial Systems, Urban Settlements, Population Health Research Platforms, Solid Earth, Marine Environment, and Characterisation.</td>
</tr>
</tbody>
</table>

Table 2: Current status of the projects in National Collections
2.4.4 Program Highlights, Issues and Breakthroughs

This year has seen ANDS entering new territory by supporting native metadata and transport protocols thus making it much easier for partners to publish metadata to the ANDS portal. Having established this functionality ANDS will develop it over the coming year with specific clients.

Through technical assistance from ANDS, all 20K+ records from Geoscience Australia now include citation metadata and improved rights metadata. This lays the groundwork for Digital Object Identifier (DOI) minting and providing metadata to Thomson Reuters Data Citation Index. This work undertaken with Geoscience Australia to significantly improve the transform from ANZLIC metadata profile to the metadata schema required by RDA, has been reused to improve records from other clients who use the ANZLIC metadata profile or the (extended) Marine Community Profile. Such clients include: CSIRO; Australian Antarctic Division; Terrestrial Ecosystem Research Network (TERN); e-Atlas; the Australian Institute of Marine Science; NSW Office of Environment and Heritage; and Integrated Marine Observing System (IMOS). In doing so, ANDS has enabled the provision of citation metadata, enabling individual researchers to receive citation for their data, and the institution to use metrics to report to their funding bodies and stakeholders.

ANDS has pursued an opportunity with the data.gov.au federal government portal to establish a bi-directional relationship so that ANDS can harvest metadata from them as well as having presence within it. ANDS undertook the work to interrogate and understand the CKAN portal technology and develop a crosswalk which could be reused with minimal customisation and a prototype for generating ANDS Registry records. The federal portal is finalising its standards and when this is complete, the team will be able to complete the first phase. ANDS intends to establish similar relationships with each of the state government portals based on this work.

A priority project within the National Collections program has been The Australian Urban Water Collection. During this period, the team initiated, or continued, engagements with partnering institutions including the Goyder Institute, the National Water Commission, the Australian Water Recycling Centre of Excellence, Water Research Australia, the CRC for Water Sensitive Cities and Smart Water Fund (Vic) to identify funded data collections that may form part of this collection. ANDS has also been involved in reviewing a Business Case for an Urban Water Research Knowledge Management portal led by the Partnership Working Group of the National Urban Water Research and Development Forum. A significant challenge for this program has been progressing the population of the Australian Urban Water Collection. There are a number of inhibitors including immature data management policies and the lack of data publication mandates by funders in this sector. In addition, many national institutions are constrained by the current fiscal environment.

The National Collections program continues to be involved in supporting feature development in RDA, including enabling increased user participation through the deployment of an annotation/ tagging facility and supporting the cross-institutional linking of thematic collections of significance. At the conclusion of this reporting period, eight new pages representing thematic views of content, were released to RDA. The themes align with the NCRIS 2011 Strategic Roadmap for Research Infrastructure and include: Integrated Biological Diversity, Astronomy, Terrestrial Systems, Urban Settlements, Population Health Research Platforms, Solid Earth, Marine Environment, and Characterisation. This is a significant breakthrough in the presentation of content, and the highlighting of national collections, in RDA.

2.4.5 Program Learnings

There were a number of learnings that emerged for the program in this reporting period. Possibly the most significant one was readiness. The development of national collections is very dependent on the desire for both the creator to publish and the demand from the research area to drive the publication.

The Urban Water distributed collection also provided a number of learnings. One that was very evident was the lack of penetration of data management policy at the funder level, thereby removing a valuable imperative to the description and management of data by funded researchers and institutions. Addressing this deficit will be critical to the ongoing success of any collection activities that take the funding body as the starting point. The experience gained from the Goyder Institute project demonstrated that attaching publications as a condition of funding can be very effective.

Across the distribution collections the development of a community of practice of those data management staff from the various institutions has been very successful. It has enabled knowledge sharing, standards development and general moral support.
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 continued building out the national infrastructure services that enable research data to be published, discovered, re-used, cited and tracked 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

RESEARCH DATA AUSTRALIA

Research Data Australia is a set of web pages (see Figure 3 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.

Research Data Australia provides an overview of Australia’s research data assets with all domains and all research organisations in scope. It provides a data-switchboard functionality to connect data seekers to data held at Australian research institutions and national data infrastructure facilities. It is also designed to emphasise the connections between disparate data and facilitate cross-disciplinary research and the addressing of grand challenges that require combined responses across domains.

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 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/

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 registry can now harvest in multiple metadata formats and transport protocols.

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.

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 the links.

Identify My Data can be accessed 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 is a member of 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.

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, 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.

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 above.

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 inter-operable 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.

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

2.5.3 Activity/Deliverables for 2013-14

Operate and develop national data discovery, data citation, and data standardisation services:

- Research Data Australia’s search algorithm was improved resulting in a more predictable and accurate search result. Complementing the improved search algorithm were a number of Search Engine Optimisation (SEO) changes to the sitemap and indexing rules for Research Data Australia.
- Implementation of record tagging functionality in RDA that provides logged in RDA users the ability to annotate Research Data Australia records with subject tags. Added tags are displayed in record view pages, indexed for searching and used in search subject facets within RDA.
Another significant improvement to the display of records in RDA was the ability to now visually identify party records that share the same identifier. Not only that similar party records are displayed systematically in the search result page, these records are also linked to each other, providing users with a more complete view of a researcher’s background, disciplines and connections.

- The ANDS Collections Registry’s Theme CMS was enhanced to optimise the display of Theme Pages in RDA. This enhancement in the Theme CMS helped in the development of eight new themes in RDA.
- To support users with unknown or unpredictable IP addresses, the ANDS Cite My Data service users have been given the option to use a shared token as an alternate (and preferred) method of authentication against the service.
- In addition, ANDS developed a Cite My Data broken link checker that runs on a quarterly basis to identify URLs associated with DOIs which no longer resolve. Details of the report are accessible from the DOI Query Tool and emailed to ANDS Services for tracking and management.
- A new version of RIF-CS was released in November 2013. RIF-CS v1.5 now supports richer relationships or linking between the registry objects and external resources such publications.
- In support of our effort to improve ANDS’ national and international presence, ANDS has completed the following activities.
  - Integration of ANDS systems with ORCID, providing Research Data Australia users the capability to import collection metadata to his/her ORCID researcher profile.
  - Support for first adopters to export from RDA to Thomson Reuters Data Citation Index.
  - Completion of a crosswalk for ARC grant activities, allowing Data Source administrators to link their records to more recent grant information and providing RDA users the ability to discover ARC grants and the associated research output more easily.

Provide solutions and support for data portal software

- Research Data Australia, Identify My Data and Cite My Data were migrated from NeCTAR’s NSP (National Servers Program) to FSP (Full Service Phase) in April 2014. The migration was in accordance with NeCTAR’s plan to host all services to the FSP infrastructure.
- Tied up with the significant improvements to ANDS’s reusable registry software, web services and widgets, ANDS launched a user-friendly Developer Toolbox (http://researchdata.ands.org.au/developers/) and Developer Community blog (http://developers.ands.org.au) to promote software reuse and encourage open source contribution from the developer community.

Support the uptake of ANDS Services

- 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. 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 v1.5 in October 2013 and has met regularly during this year to deliberate on a new version for October 2014. 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 2010-14 with historic figures from the previous years:

### Public System

<table>
<thead>
<tr>
<th>Service</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</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>9,667</td>
</tr>
<tr>
<td>Trusted SW Agreement</td>
<td>10</td>
<td>11</td>
<td>14</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

### DOI Service (Cite My Data): 1

| DOIs Minted (production) | -        | -        | 29       | 1,712    | 2,405    |
| Registered Publisher Agents | -        | -        | 12       | 22       | 29       |
| Publisher Agents in Production | -        | -        | 4        | 10       | 12       |

### Registry

<table>
<thead>
<tr>
<th>Data Source Admins</th>
<th>-</th>
<th>46</th>
<th>185</th>
<th>224</th>
<th>233</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider Org</td>
<td>12</td>
<td>21</td>
<td>59</td>
<td>68</td>
<td>71</td>
</tr>
<tr>
<td>Total data source feeds</td>
<td>12</td>
<td>45</td>
<td>122</td>
<td>150</td>
<td>157</td>
</tr>
<tr>
<td>Direct</td>
<td>11</td>
<td>30</td>
<td>60</td>
<td>65</td>
<td>56</td>
</tr>
<tr>
<td>Automatic</td>
<td>1</td>
<td>15</td>
<td>62</td>
<td>85</td>
<td>101</td>
</tr>
<tr>
<td>Publish my Data</td>
<td>51</td>
<td>63</td>
<td>60</td>
<td>46</td>
<td>46</td>
</tr>
</tbody>
</table>

### Research Data Australia:

<table>
<thead>
<tr>
<th>Collections</th>
<th>2,906</th>
<th>26,746</th>
<th>72,990</th>
<th>106,976</th>
<th>164,168</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parties</td>
<td>894</td>
<td>1,960</td>
<td>5,473</td>
<td>14,681</td>
<td>25,616</td>
</tr>
<tr>
<td>Services</td>
<td>3</td>
<td>30</td>
<td>84</td>
<td>141</td>
<td>183</td>
</tr>
<tr>
<td>Activities</td>
<td>690</td>
<td>857</td>
<td>27,476</td>
<td>35,555</td>
<td>40,697</td>
</tr>
</tbody>
</table>

### TRIAL SYSTEM 4

<table>
<thead>
<tr>
<th>Handle Service (Test):</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handles minted</td>
<td>-</td>
<td>15,399</td>
<td>32,937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trusted SW Agreement</td>
<td>-</td>
<td>30</td>
<td>55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Sandbox:

| Data Source Admins     | 91       | 128      | 137      |          |          |
| Provider Org           | 60       | 105      | 109      |          |          |
| Total Records          | 33,258   | 85,881   | 69,342   |          |          |
| Collections            | 3,753    | 32,416   | 40,233   |          |          |
| Parties                | 2,392    | 26,270   | 7,519    |          |          |
| Services               | 29       | 44       | 114      |          |          |
| Activities             | 27,084   | 27,151   | 21,476   |          |          |

Table 3: Service usage reports for 2010-14

### NOTES:

1. The DOI Service, Cite My Data, was launched last December 2011. This figure shows DOIs minted since its release.
2. Page Views - Research Data Australia page views including hits from crawlers and robots (data gathered from the local server logs). Google Analytics tools was used to gather page view reports from July 2010. Starting July 2013, the Google analytics report shows visits and visitor as added information.
3. Google analytics Reports: Pageviews (Hits) vs. Unique Pageviews: A pageview is defined as a view of a page on your site that is being tracked by the Analytics tracking code. If a visitor clicks reload after reaching the page, this is counted as an additional pageview. If a user navigates to a different page and then returns to the original page, a second pageview is recorded as well. A unique pageview aggregates pageviews that are generated by the same user during the same session. A unique pageview represents the number of sessions during which that page was viewed one or more times. - Unique Visitors: The number of users that viewed or interacted with RDA content within a specific date range. Visits vs Visitors: Visits represent the number of individual sessions initiated by all the visitors to your site. If a user is inactive on your site for 30 minutes or more, any future activity is attributed to a new session. Users that leave your site and return within 30 minutes are counted as part of the original session. The initial session by a user during any given date range is considered to be an additional visit and an additional visitor. Any future sessions from the same user during the selected time period are counted as additional visits, but not as additional visitors.
4. 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).
**ANDS Handle Service Report - Handles Minted**

**Annual Summary - 2010-2014**

<table>
<thead>
<tr>
<th>Year</th>
<th>PUBLIC SYSTEM</th>
<th>TRIAL SYSTEM 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>4,279</td>
<td>-</td>
</tr>
<tr>
<td>2011</td>
<td>6,323</td>
<td>15,399</td>
</tr>
<tr>
<td>2012</td>
<td>6,830</td>
<td>32,937</td>
</tr>
<tr>
<td>2013</td>
<td>9,015</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>9,667</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 6: Handles minted**

**ANDS Handle Service Report - Trusted SW Agreement**

**Annual Summary - 2010-2014**

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

**Figure 7: Trusted SW Agreements**

**ANDS Registry Report - Provider Organisations**

**Annual Provider Summary - 2010-2014**

<table>
<thead>
<tr>
<th>Year</th>
<th>PUBLIC SYSTEM</th>
<th>TRIAL SYSTEM 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td>2011</td>
<td>21</td>
<td>105</td>
</tr>
<tr>
<td>2012</td>
<td>59</td>
<td>109</td>
</tr>
<tr>
<td>2013</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>71</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 8: Provider Organisations**
ANDS Registry Report - Data Provider Accounts
Annual Provider Summary - 2010-2014

<table>
<thead>
<tr>
<th>Year</th>
<th>PUBLIC SYSTEM</th>
<th>TRIAL SYSTEM 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>-</td>
<td>91</td>
</tr>
<tr>
<td>2011</td>
<td>46</td>
<td>128</td>
</tr>
<tr>
<td>2012</td>
<td>185</td>
<td>137</td>
</tr>
<tr>
<td>2013</td>
<td>224</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>233</td>
<td></td>
</tr>
</tbody>
</table>

Figure 9: Data Provider Accounts

Note: Data Provider Accounts are administrator representatives from organisations who manage feeds of information from their organisation to the ANDS registry.

ANDS Registry Report
Annual Total Records Summary - 2010-2014

<table>
<thead>
<tr>
<th>Year</th>
<th>PUBLIC SYSTEM</th>
<th>TRIAL SYSTEM 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>2,906</td>
<td>33,258</td>
</tr>
<tr>
<td>2011</td>
<td>26,746</td>
<td>85,881</td>
</tr>
<tr>
<td>2012</td>
<td>72,990</td>
<td>69,342</td>
</tr>
<tr>
<td>2013</td>
<td>106,976</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>164,168</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10: ANDS Registry Report
2.5.4 Program Highlights, Issues and Breakthroughs

The year 2013-14 saw some significant highlights in the ANDS Online Services suite. The finalisation of RDA theme pages (https://researchdata.ands.org.au/themes) represents a new possibility for exploring Australia’s research data assets. The service has launched with eight new themes and will provide a subject-based platform upon which to build into the future.

The new ANDS harvester service introduces a new paradigm in metadata harvesting which supports (potentially) any metadata format delivered using any transport protocol. This flexibility will provide a facilitating and enabling infrastructure in building the coverage of Research Data Australia.

ANDS systems are integrated with ORCID which allows researchers to link any ANDS-registered dataset to their ORCID profile with one click.

ANDS has produced an export to Thomson Reuters Data Citation Index which allows ANDS-registered data to be tracked for citations in the scholarly journals.

2.5.5 Program Learning

Until this point the ANDS data portal Research Data Australia has been mainly focused on the needs of data providers (data archives, repositories, research organisations) so as to build up a critical mass of data collection descriptions. After a few years of operation the portal’s functioning can now be fine-tuned to cater better for the needs of data seekers. As part of the ongoing maintenance of the site ANDS is compiling site analytics, behaviour patterns, user feedback, environmental scans and researcher interviews to enable a more user-focused experience on Research Data Australia.

![Annual Page Views Report](image-url)
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).
- 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 co-ordinated 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 was a series of compelling demonstrations of the value of having data available for reuse. These were required to:

- 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.
- engaged with national research capabilities.

2.6.2 Outline of projects

Consistent with these criteria for demonstrations of value, ANDS 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. The complete set of projects includes the following:

- **TROPICAL DATA HUB AND TOOLS (AP01/AP02) – $125K**
  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 created a set of complementary tools consisting of server side data manipulations and web accessible interfaces that are designed to process data stored in the TDH and related repositories. The tools enabled the creation of synthetic data products to solve specific scientific and socio-economic questions related to land use and conservation in the tropics. This project is an exemplar of the power of the ANDS data centric research model and also demonstrated the value of tropical data integration to research and policy/planning.

- **EDGAR – JAMES COOK UNIVERSITY (AP03/AP30) – $250K**
  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.

The Edgar project was closely coordinated with a project that reuses data from the Atlas of Living Australia to integrate species distribution data, locality data with climate change scenarios in an integrated fashion which facilitates the modelling of current and future species distributions based on climate scenarios. This project (AP30) essentially provided the back-end functionality for Edgar.

ADAPTATIONS AND IMPACTS DOWNSCALING TOOL (AP04) – $492K
Climate impacts research depends on the timely provision of data sets that describe how climate will change across a region. These data sets are created by climate scientists in a discipline specific format via a downscaling procedure designed by climate scientists for climate scientists. This ANDS project has built tools that identify suitable downscaled data sets and then read the contents of those data sets. A dialog with four key impacts research groups was established to define their data needs for their research. The tools were specifically developed 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. There have been multiple spin-offs resulting from this project: tools usable by the climate community to interrogate climate data sets hosted at an RDSI node, the generation of key data sets required by future downscaling projects, and the enabling of other climate impacts groups – beyond the four identified – to undertake their research by providing the tools they need to engage with the climate science community.

MARINE VIDEO (AP05) – $255K
This project provided funding to support the development of a database framework applicable to marine imagery that facilitated the timely delivery of standardized, quantitative estimates of ecologically relevant indicators (such as absence/presence, percent cover, abundance and distribution of benthic organisms and associated substrates) through the analysis of visual data being produced by AUV, BRUVS, ROV and UTV systems. The resulting software also underpinned a very successful citizen science activity, working with the ABC to develop the http://exploretesthesaloor.net.au tool. This tool provided an interface for citizen scientists to help with the analysis of imagery collected by the AUV. During the course of 10 days of Science Week 2013 the project was promoted on the ABC and through online resources such as Facebook. During this period, nearly 9000 participants signed up to use the system and labelled nearly 300,000 images.

EXTREME WEATHER AND HEALTH IMPACTS – SWISH (AP07) – $200K
The 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 for the kind of data sourcing, manipulation and analysis that is required for greater efficiency in many epidemiological research areas. The result was the ability to publish specific workflows online that integrate datasets and produce an analysis assessing and projecting the health impacts of extreme weather events. The enhanced infrastructure will eventually provide support for a diversity of epidemiology researchers (e.g., bio surveillance, wildlife health, and emerging infectious diseases).

VALIDATION OF GENOMES AND TRANSCRIPTOMES WITH PROTEOMIC DATA (AP11) – $246K
High throughput (“next-generation”) gene sequencing technology has become widespread in its use in recent years due to a significant decrease in cost. When used to sequence the DNA (the ‘genome’) of a new species, this information can be used to predict which genes (or variants of these genes) are present in the species. However, validation of this prediction requires comparison against another biological output such as 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. The tool developed in this project greatly facilitated these analyses. The project outcomes were of immediate applicability to (i) users of the Ramaciotti Centre for Gene Function Analysis and the Bioanalytical Mass Spectrometry Facility at UNSW who wish to co-analyse next-gen sequencing data with proteomic data, and (ii) the framework dataset project of Bioplatforms Australia, especially for wine yeast, soil-dwelling microbes and wheat pathogenomics projects.

POSITIVE PLACES : SPATIAL ANALYSIS OF PUBLIC OPEN SPACE (AP12) – $250K
Public open space [POS], including parks, confer a number of physical, psychological and social health benefits for individuals and the community. However there is a paucity of data on provision and spatial distribution of POS, and particularly adequate data to examine the provision and quality of POS by social disadvantage.
The key benefit to researchers and users of Soils to Satellites is being able to go to one place and visualise patterns and relationships across different data, that if the user had looked at the individual databases they came from, they just wouldn’t be able to see.

MARTIN PULLAN, TERN ECO-INFORMATICS PROJECT MANAGER, SOILS TO SATELLITES

This project 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 supported two new opportunities. In research it enabled investigations into the provision, access, and quality of POS and contribute to research efforts to understand the direct and indirect health benefits afforded by POS. For urban planning practice it enabled better planning of the land allocation and positioning of POS, allocation of POS amenities relative to existing large and local scale services; and the modelling of future needs of POS according to forecasted and hypothetically modelled demographic changes.

SMART’S MULTI-UTILITY DASHBOARD - INFRASTRUCTURE ANALYTICS FOR INTEGRATIVE RESEARCH (AP14) – $215K

The SMART Infrastructure Facility developed a ‘multi-utility dashboard’ that offers infrastructure analytics based on data provided by public agencies and private operators. The online dashboard allows analysts to develop new insights into spatial, technical, social and economic issues associated with regional and urban infrastructure development. The multi-utility dashboard acts as ‘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 powerful infrastructure analytics. The multi-utility dashboard offers a flexible collaborative platform to researchers, business analysts and local planners.

SOILS TO SATELLITES (AP15) $500K

The project allows for the identification, exploration and comparison of similarities and differences across various environmental sites across Australia. The tool that was developed integrated 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 these sites. The sites are 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 this data indicates the relatedness of the communities and turn over in community composition along environmental gradients. The genetic differentiation between populations of species amongst sites provides an indication of the level of gene flow, and therefore dispersal, between sites and across the landscape (tempered by environmental selection pressures and other drivers). Such correlations are important for understanding the likely redistribution dynamics and resilience of communities and species under predicted climate change.

BRAIN MAPPING NATIONAL RESOURCE – TISSUESTACK (AP16) $399K

The Centre of Advanced Imaging at UQ 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 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.

MULTIMODAL KIDNEY IMAGING DATA INTEGRATION (AP19) $200K

This project developed a tool that will inform future potential non-invasive, imaging based procedures for monitoring kidney disease progression. The project built a system 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 in the future on living subjects). The work was carried out using data collected from rat specimens and this project resulted in multi-modal image data of rat kidneys to be integrated to produce an analysis resource that will help elucidate biological and molecular function and potentially advance diagnosis and treatment of human renal diseases.
The Founders and Survivors Project (www.foundersandsurvivors.org) has brought together a number of research data sets created from records relating to the 73,000 convicts transported to Tasmania in the nineteenth century and their descendants to create a population database of national and international significance for historical, demographic and population health researchers. This project built on that work by developing open-source software infrastructure based on the Yggdrasil database that enables 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.

The Australian Pancreatic Cancer Genome Initiative (APGI) is delivering genomic DNA and RNA sequence data associated with pancreatic tumour samples as part of the International Cancer Genome Consortium (ICGC). Their aim is to profile tumour and normal tissue of approximately 375 pancreatic cancer patients. For this, the APGI uses the best clinical material available, with well-characterised and accurately annotated clinico-pathological, treatment and outcome data acquired prospectively. However, the effective analysis and re-use of these DNA/RNA datasets of international importance by research biologists and clinicians is currently limited by their ability to access and to use computational and data infrastructure. This project has developed software that enables researchers and clinicians to access, integrate and transform unique and valuable cancer
The Proteomic-Genomic Nexus, developed at the Systems Biology Initiative at the University of New South Wales, has already been used in several interesting projects, including validating the genome of a strain of bacteria associated with Crohn’s disease and tracing the changes of proteins associated with the differentiation of bone cells.

The 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 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. It developed a software infrastructure allowing 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. The project deliverables are of strategic value to climate and carbon policy makers in Australia, and of great utility for benchmarking ecosystem models in Earth system science.

HUMAN PROTEOME BROWSER (AP32) $200K
10 years ago the first draft of the human genome was published and currently the estimate of the number of human protein coding genes is around 21,000. While the amount of information and complexity of the genome is large, the level of complexity increases considerably when one considers the functional output of the genome – i.e. proteins. Current estimates of the number of proteins encoded by the 21,000 genes stands in the region of ~2,000,000. HUPO is an international scientific organisation representing and promoting proteomics through international cooperation and collaborations, and is directing an international effort to identify all human proteins. In this scheme, Australia has been assigned Chromosome 7, as this chromosome contains various genetic markers associated with diseases relevant to the Australian population. In this global effort, much experimental work will be carried out in a methodical manner to identify all predicted proteins, followed by cataloguing and subsequent analysis of the data. The "Human Proteome Browser" developed in this project will integrate a wealth of currently publicly accessible, yet disparate protein evidence together into a simple yet powerful interface. This will function as a portal to the human proteome for the global research communities and greatly expedite the HUPO international effort. The tool is currently being evaluated by the human proteome activity for wider deployment.

DATA TRANSFORMATION AND MODEL CALIBRATION SYSTEM (AP29) – $200K
Carbon and nitrogen dynamics are important factors in managing and understanding Australian ecosystems. Historic data are key for running predictive models of Carbon and nitrogen dynamics in the ecosystem. These models are generic and need to be calibrated and validated for Australian conditions. A working group sponsored by the Australian Centre for Ecological Analysis and Synthesis (www.aceas.org.au) is collecting and nitrogen data from a number of Australian sites in a standard format. These data will subsequently be fed to a number of models as training data to improve the accuracy of the models. The performance of the models will then be compared to test data to gauge their adequacy and the validity of their outputs. This project has built a data transformation and model calibration system that automatically or semi-automatically extracts and transforms historic data from multiple Australian sites and uses them to calibrate and validate the predictive power of multiple carbon and nitrogen dynamics models. The software captures the process in a scientific workflow which will form the blueprint for future similar activities to calibrate and compare ecosystem models using Australian data.

AURIN & ANDS – NORTH WEST METROPOLITAN REGION OF MELBOURNE DATA ACCESS, INTEGRATION AND INTERROGATION AND DEMONSTRATOR PROJECTS (AP31) – $275K
This project has facilitated access to a myriad of data sets for the Melbourne North West corridor, providing access to a number of datasets to researchers via the AURIN portal. This unprecedented data access will enable world-class research that will be focused toward addressing the key policy issues in the North West region of Melbourne, as identified by the North and West Metropolitan Regional Management Forum. The aim of the project was to demonstrate the benefit of providing open access of government datasets to researchers, planners and policy makers in dealing with problems of space, place, and liveability. The value of the project was demonstrated through four demonstrator projects which covered four of the most pressing issues facing the North West Melbourne Region: built environment and health, housing affordability, economic productivity, and transport and sustainability.
SEQUITOR: A DEMONSTRATION INTEGRATED COASTAL KNOWLEDGE PLATFORM (AP34) – $300K

Management across the catchment-coast-ocean continuum remains one of the greatest Natural Resource Management challenges still to be addressed in Australia. This project has developed data integration and visualisation infrastructure for a proof-of-concept Coastal Knowledge platform for South-east Queensland known as SEQUITOR. SEQUITOR 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 project has exposed and integrated disparate datasets from a number of organisations in South-east Queensland, visualised and analysed disparate data in a publically accessible web environment, and assimilated data from observational networks to better constrain catchment and receiving environment models. The project has also been extended to visualise the recent Logan River floods, combining the kinds of data listed above with crowd-sourced photos, videos and textual narratives. The result is likely to be adopted by the Logan Council as part of their process of engaging with the community that they serve.

DECISION SUPPORT TOOLKIT FOR CLIMATE RESILIENT SEAPORTS (AP35) – $253K

This project 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 which enables Australian port authorities to make better informed climate risk management decisions. The project needed to do three discrete things. The first was the sourcing, refinement, and standardisation of multiple data sets needed for context specific climate change adaptation decision-making. Once acquired, both primary and transformed data (as produced by new extensible models being developed for the NCCARF project) needed to be integrated within a suitable data management framework. The last step was the development of innovative software that provides an interactive interface (developed in close consultation with port authorities) for considering future climate change impacts, the implications for organisational risk management strategies, and the identification of possible adaptation options. The final product is an online decision support toolkit that can be readily used by port authorities both in Australia, and potentially internationally.

2.6.3 Activity/Deliverables for 2013-2014

During the reporting period the following projects completed and met all the contracted deliverables:

- Adaptations and Impacts Climate Model Downscaling Tool.
- Extreme Weather Impacts and Health Events (SWISH).
- University of Wollongong SMART Utility Dashboard.
- Validation of genomes and transcriptomes with proteomic data.
- Multi-scale Kidney Imaging.
- Founders and Survivors: Genealogical Connections.
- Brain and Mind Research Institute.
- Climate Change Adaptation Information Hub – Terranova.
- Marine Video.
- Marine Virtual Laboratory Information System.
- Primary production in space and time (ePISAT).
- Data Transformation and Model Calibration System for Carbon and Nitrogen Dynamics in Australian Ecosystems.
- Soils 2 Satellites.
- Spatial Analysis of Public Open Space.
- AURIN & ANDS – NW Metro Region Demonstrators.
- Catchment to Coast (SEQUITOR).

While all of the projects completed successfully, many of them ran late. As a number of them had very aggressive timelines, 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 that were still underway during the reporting period continued to blog about their activities, and the results can be viewed at [http://andsapps.blogspot.com.au/p/project-feed.html](http://andsapps.blogspot.com.au/p/project-feed.html).
2.6.4 Program Highlights, Issues and Breakthroughs

A particularly pleasing aspect of the projects this year is the way that the code developed in one project has been picked up by successor projects:

- SEQUITOR has been built on by TERN for the Logan Flood Visualisation [http://loganvis.meteor.com/] project, which in turn may be picked up by AURIN.
- Project Edgar code is being used for one of the JCU MODC collections.
- EPISAT code is being used for the Macquarie MODC collection.
- MARVLS is being augmented for the Marine eRIC project.

Other highlights were 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 developed strong connections (both personal and technological) with each other.

One of the issues confronting the Applications program this year was the time taken to wrap up the last tranche of projects. As projects got close to completion, the project staff were often deployed onto something new, meaning that they found it difficult to make the time to wrap up things like code deposit and final reports.

2.6.5 Program Learning

Engaging in discussions with research champions about barriers they face to bringing data together was very instructive. A number of them commented that the kind of funding that ANDS provided was 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 was able to fund the latter but not the former. Together, a rich range of possibilities became available. It would be good to identify a way to continue this sort of activity.

We’re fortunate in that the Australian National Data Service, otherwise known as ANDS, saw this as an important imperative and were in a position to fund this work. Through their support we built a software platform to do these complex cross comparisons. These tools are designed to be easy to use, and we expect they’ll help many researchers to really co-explore genomic, proteomic and transriptomic data. We’ve also established a means by which to feed genomics and proteomics metadata through to the data fabric that is established and run by the Australian National Data Service. This will help make these large volumes of data which are being produced more discoverable.
2.7 International Infrastructure

2.7.1 Overview of program

The aim of this program was (and 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 therefore chose to pursue international opportunities in order to:

- influence developing international infrastructure initiatives to support the ANDS mission by adopting technologies and approaches compatible with those adopted by ANDS.
- 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.
- ensuring that Australian researchers have access to and can 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.
- 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, consistent with the overall ANDS approach, included an ongoing emphasis on the importance of richly described and connected research data collections, and a focus on the significance of research institutions in managing research data.

2.7.2 Outline of projects

This activity did not involve the funding of any projects, but was based on contribution of effort.

2.7.3 Activity/Deliverables for 2013-2014

ANDS staff remain closely involved in the governance of the Research Data Alliance (RDA). Dr Ross Wilkinson is still a member of the Council (the main governing body), Dr Andrew Treloar is co-chair of the Technical Advisory Board, and Dr Stefanie Kethers is a key member of the Secretariat.

The Australian Commonwealth government provided additional funding to ANDS to take part in RDA activities, and to support the Third Plenary. This took place in March of 2014, and attracted over 500 participants.

Beyond its RDA activities, ANDS staff took place in a range of other international activities:

- ANDS was 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 continues to contribute to the direction and growth of the DataCite Consortium, an organisation established to provide digital object identifiers for data.
- ANDS continued to work directly with national counterparts in the UK, Netherlands, Finland, Germany and Denmark.
2.7.4 Program Highlights, Issues and Breakthroughs

The main highlight over this period was the RD-A Third Plenary in Dublin, on March 26 to 28. Over 500 people attended this event which was very much a joint Australian-Irish affair. One of the keynote speeches was delivered by Australia’s Chief Scientist and the Australian Ambassador to Ireland took part in the opening ceremonies. Thirteen Australian delegates attended. The numbers attending were significantly up on the previous event, and the level of energy was very high. There has also been a steady increase in the membership of the Research Data Alliance (RD-A) as a whole.

In a non-Research Data Alliance context, one highlight was the successful completion of the EU-funded ODIN project, in which ANDS is a non-funded partner. Another is the ongoing engagement with Thomson Reuters over the Data Citation Index and determining the best way to provide a feed from Australian data providers. ANDS has also been working with its partners to help them prepare their records for passing across to Thomson Reuters.

An ongoing 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 required significant overseas travel, although this was 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.

ANDS has played a seminal role in providing research data environments for Australia’s researchers. Its efforts are very well recognised internationally through its links with the Research Data Alliance.

ANDS STEERING COMMITTEE CHAIR DR RON SANDLAND
SHARE ISSUE 17, OCTOBER 2013
2.8 Project Office

2.8.1 Overview of program

This program is designed to ensure the effective and efficient delivery of the ANDS project with proper reporting of outcomes and compliance with all governance requirements. It has also adapted to support the delivery of all ANDS outcomes with the goal of delivering operational excellence. The workplace functions that sit under the project office are finance, communications, contract management, governance and reporting, internal IT systems, office administration and general operational support.

2.8.2 Outline of projects

Taking 'Excellence in operational management' as its goal, the program has been continually focused on streamlining processes and increasing the cohesion of the ANDS group for a more effective overall delivery of internal and external support. This has been achieved by reviewing existing processes and where relevant introducing new lightweight processes that allow for greater cross organisational flexibility.

2.8.3 Activity/Deliverables for 2013-14

Over the last reporting period the activity of the Project Office team has been focused on supporting the ANDS team during what has been a challenging period. This period has seen an important transition in the life of the ANDS project as the remaining 60 projects funded under NCRIS and EIF concluded. During the reporting period, the office:

• Provided a high level of administration and project support to ensure that all relevant programs funded under NCRIS and EIF were concluded with funds exhausted by the acquittal deadline.
• Continued to drive a cohesive communications message to the sector using share, our e-newsletter andsUP and the social media platforms that we engage with.
• Provided ongoing training and support to non-administrative staff to reduce the amount of central administration effort required.
2.9 Promotion

ANDS has undertaken a large number of promotional activities during the period July 2013 to the end of June 2014. These include the following.

2.9.1 Presentations/attendance at Conferences & Workshops

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:

- ODIN 1st year conference and code sprint, Geneva.
- eResearch Australasia, Brisbane.
- Research Profiling Conference, University of Melbourne.
- RD-A plenary 3, Dublin, Ireland.
- ODIN meeting, British Library, London.
- JISC/IDCC workshop, San Francisco.
- Open Repository Conference, Helsinki.
- Research Data Immersive Informatics course: University of Melbourne.
- Visualisation launch CAVE2, Monash University.
- Doing Data Better, Miniconference, University of Melbourne.
- RMIT Big Data Symposium, Good Data, Bad Data.
- Research Community Day, University of Melbourne.
- Copyright Council Training, Melbourne.
- Apps4NSW, NSW Government, Sydney.
- RDA Plenary 2, Washington.
- Digital Curator Workshop, Sydney.
- International Digital Curation Conference, San Francisco.
- CSIRO eResearch Conference, Melbourne.
- Locate 14 Spatial Conference, Canberra.
- Committee on Publication Ethics.
- IEEE eScience’13, Beijing.
- CAUL Research Advisory Committee meetings, Brisbane.
- Thomson Reuters CONVERIS Roadshow, University of Queensland.
- PLOS and Open Access seminar with Dr Virginia Barbour, University of Queensland.
- DataCite Summer Meeting, Washington DC.
- Belmont Forum Steering Committee, Windsor.
- Third EU-AU Research Infrastructure Workshop, Canberra.
- Big and Complex Data Workshop, Melbourne.
- TERN and eResearch Infrastructure Workshop, Canberra.
- PARADISEC Conference, Melbourne.
- SKA Big Data Workshop.
- International Conference on Research Infrastructure, Athens.
- eResearch 2020 Workshop, Wellington, NZ.
- Belmont Forum Steering Committee, Vienna.
- TERN ACEAS Grand Synthesis Workshop, University of Queensland.
- TERN Flood Visualisation project stakeholders Workshop, University of Queensland.
- NeCTAR eResearch Projects Workshop, Melbourne.
- Australian Phenomics Workshop Strategic Planning Workshop, Adelaide.
- ARMS conference, Adelaide.
- Polar Data Forum, Tokyo.
- Open Access Conference, Brisbane.
- 2nd Research Support Community day, Melbourne.
- AeRO National Forum, Perth.
- AuScope Symposium, Canberra.
- Urban Water R&D forum, Canberra.
- TERN IIDDG meeting, Canberra.
- Bioplatforms Australian and eResearch Coordination meeting, Canberra.
- QUT Open access and Research Conference, Brisbane.
- RD-A Council meeting, London.
- Meeting of the Data Infrastructures working group, London.
- Open Data Research Portal workshop, Gold Coast.
- AusAID Symposium, Brisbane.
- AU Embassy visit, Brussels.
- 11th Australasian Data Mining Conference, Canberra.
- Data Management and Sharing in Government, Canberra.
- CSIRO eResearch Conference, Melbourne.
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 hosted a variety of free virtual events in the period 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.

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.

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 three issues in the period have focused on Australia’s role in global research data, celebrating the completion of ANDS funded projects and the value of research data.

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), Australian Research Management Society and Public Library of Science. 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.

The number of @andsdata followers has increased steadily over the last four years:

- 79 followers on 16 November 2011 (two weeks after the launch of @andsdata).
- About 450 followers in April 2013.
- 864 followers on 12 May 2014.
- 978 followers on 14 August 2014.

@andsdata, on average, currently gains 1 new follower per day, from Australia or overseas.

@andsdata followers are currently from the following countries/regions:

<table>
<thead>
<tr>
<th>Country</th>
<th>Followers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>52%</td>
</tr>
<tr>
<td>Europe</td>
<td>25%</td>
</tr>
<tr>
<td>US</td>
<td>16%</td>
</tr>
<tr>
<td>New Zealand</td>
<td>2%</td>
</tr>
<tr>
<td>Africa</td>
<td>1%</td>
</tr>
<tr>
<td>Middle East</td>
<td>1%</td>
</tr>
<tr>
<td>Asia</td>
<td>2%</td>
</tr>
<tr>
<td>South America</td>
<td>1%</td>
</tr>
</tbody>
</table>

Those who have mentioned @andsdata positively in their tweets include The Dataverse Project, Harvard University, and the British Library Science Team.
Figure 12: Twitter report outlining global audience
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 13: Youtube report outlining overview of performance and engagement

Figure 14: Youtube report outlining overview of demographics
ANDS maintains a Risk Register. The risk assessment methodology, adapted from the Australian Risk Management Standard AS/NZS 31000:2009, 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 February 2014. ANDS assessed the residual risk level of all 11 existing 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 three changes from the previous review in June 2013.

Risk 1 had reduced from a High Risk to a Low risk due to recent changes around ARC funding. Risk 3 had reduced from a high risk to a medium risk as the number of outstanding has dramatically reduced with only 16 remaining. Also under the NCRIS 2013 plan we will have a much small amount of external contracts. A new Risk was added, Risk 12 - That ANDS is most effectively reporting under all the various programs.
3. Progress against milestones

3.1 Data Capture Infrastructure

These are the milestones for the Institutional Engagement Program in the 2013-14 Annual Business Plan. Progress is reported against the NCRIS Seeding the Commons Program, the EIF Data Capture Program, the EIF Metadata Stores Program, and the ARDC Application Program.

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>13Q3</td>
<td>Five institutional agreements on work programs for their research data assets</td>
<td>In the course of 2013-2014 ANDS started conversations with the universities in the Major Open Data Collections program. This is reported on in the NCRIS report. Some of the projects took longer to complete. ANDS has invested effort in supporting institutions throughout 2013-14 in completing these projects.</td>
</tr>
<tr>
<td></td>
<td>Successful conclusion of all Data Capture, Seeding the Commons, Metadata Stores and Applications projects</td>
<td>See above</td>
</tr>
<tr>
<td>13Q4</td>
<td>Further five institutional agreements on work programs for their research data assets Identification of some institutional assets</td>
<td>See above</td>
</tr>
<tr>
<td>14Q1</td>
<td>Further five institutional agreements on work programs for their research data assets Identification of some institutional assets Concluded agreed work on research data assets with initial five institutions</td>
<td>See above</td>
</tr>
<tr>
<td>14Q2</td>
<td>Further five institutional agreements on work programs for their research data assets Identification of some institutional assets Concluded agreed work on research data assets with further five institutions</td>
<td>See above</td>
</tr>
</tbody>
</table>
3.2 Metadata Store Infrastructure

These are the milestones for the Institutional Engagement Program in the 2013-14 Annual Business Plan. Progress is reported against the NCRIS Seeding the Commons Program, the EIF Data Capture Program, the EIF Metadata Stores Program, and the ARDC Application Program.

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
</table>
| 13Q3           | Five institutional agreements on work programs for their research data assets  
Successful conclusion of all Data Capture, Seeding the Commons, Metadata Stores and Applications projects | In the course of 2013-2014 ANDS started conversations with the universities in the Major Open Data Collections program. This is reported on in the NCRIS report  
Some of the projects took longer to complete. ANDS has invested effort in supporting institutions throughout 2013-14 in completing these projects |
| 13Q4           | Further five institutional agreements on work programs for their research data assets  
Identification of some institutional assets | See above |
| 14Q1           | Further five institutional agreements on work programs for their research data assets  
Identification of some institutional assets  
Concluded agreed work on research data assets with initial five institutions | See above |
| 14Q2           | Further five institutional agreements on work programs for their research data assets. Identification of some institutional assets. Concluded agreed work on research data assets with further five institutions | See above |
3.3 Public Sector Data Infrastructure

These are the milestones for the National Collections Program in the 2013-14 Annual Business Plan. Progress is reported against the NCRIS National Collections Program, and the EIF Public Sector Data Program.

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>13Q3</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
<td>Engagement with institutions:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• TERN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Australian Water Recycling Centre of Excellence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Smart Water Fund</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• National Water Commission</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Water Research Australia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• CRC for Water Sensitive Cities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bureau of Meteorology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Water Services Association of Australia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• NSW Office of Environment &amp; Heritage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• National Computational Infrastructure [NCI]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data.gov portals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Geoscience Australia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Commonwealth Department of the Environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• e-Atlas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Australian Institute of Marine Science</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Population Health Research Network</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Public Records Office of Victoria</td>
</tr>
<tr>
<td></td>
<td>Three topic pages and two themed distributed collections published</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Engagement with institutions:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• TERN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Australian Water Recycling Centre of Excellence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Smart Water Fund</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• National Water Commission</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Water Research Australia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• CRC for Water Sensitive Cities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bureau of Meteorology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Water Services Association of Australia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• NSW Office of Environment &amp; Heritage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• National Computational Infrastructure [NCI]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data.gov portals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Geoscience Australia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Commonwealth Department of the Environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• e-Atlas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Australian Institute of Marine Science</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Population Health Research Network</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Public Records Office of Victoria</td>
</tr>
<tr>
<td></td>
<td>Services associated with themed distributed collections available via RDA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further two themes commenced for distributed collections</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further three topic pages published</td>
<td></td>
</tr>
<tr>
<td>13Q4</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
<td>Collections published:</td>
</tr>
<tr>
<td></td>
<td>Services associated with themed distributed collections available via RDA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further two themes commenced for distributed collections</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further three topic pages published</td>
<td></td>
</tr>
<tr>
<td>14Q1</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
<td>Collections published:</td>
</tr>
<tr>
<td></td>
<td>Collections published with initial three institutions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Services associated with themed distributed collections available via RDA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further three topic pages published</td>
<td></td>
</tr>
<tr>
<td>14Q2</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
<td>Collections published:</td>
</tr>
<tr>
<td></td>
<td>Continued engagement with next three institutions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Collections published with next three institutions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further three topic pages published</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Topic (theme) pages published (8) with associated services available via RDA:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Marine Environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Terrestrial Systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Solid Earth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Astronomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Integrated Biological Discovery</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Characterisation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Urban Settlements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population Health Research Platforms</td>
</tr>
</tbody>
</table>
## 3.4 ARDC Core Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>2013-14 Milestones</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>13Q4</strong></td>
<td>1. Citation proof of concept projects: mid project reports received</td>
<td>1. Citation mid-term project reports received</td>
</tr>
<tr>
<td></td>
<td>2. Community support webinars held: Geospatial, Data Citation, Identifiers, Vocabularies</td>
<td>2. Community support webinars held for Geospatial, Data Citation, Identifiers, Vocabularies. See: <a href="http://ands.org.au/presentations/audio-video.html">http://ands.org.au/presentations/audio-video.html</a></td>
</tr>
<tr>
<td></td>
<td>3. Briefings provided to DIICCSRTE, ARMS CAUL on national data policy agenda</td>
<td>3. Policy briefings provided to DIICCSRTE (and successors) ARMS and CAUL</td>
</tr>
<tr>
<td></td>
<td>5. ORCID systems integration complete</td>
<td>5. ORCID systems integration complete. See: <a href="https://www.youtube.com/watch?v=KyUNFECg9x4">https://www.youtube.com/watch?v=KyUNFECg9x4</a></td>
</tr>
<tr>
<td><strong>14Q1</strong></td>
<td>1. Community support events held: National Collections</td>
<td>1. National collections community support events held: Australian Antarctic Data Archive workshop</td>
</tr>
<tr>
<td></td>
<td>2. Briefings and submissions provided to funding agencies on national data policy agenda</td>
<td>2. Briefings and submissions provided to RDIC and ARCom</td>
</tr>
<tr>
<td><strong>13Q2</strong></td>
<td>1. Citation proof of concepts: final reports received</td>
<td>1. Citation proof of concept projects completed. See: <a href="https://projects.ands.org.au/id/FC012">https://projects.ands.org.au/id/FC012</a> and <a href="https://projects.ands.org.au/id/FC013">https://projects.ands.org.au/id/FC013</a> and <a href="https://projects.ands.org.au/id/FC014">https://projects.ands.org.au/id/FC014</a></td>
</tr>
<tr>
<td></td>
<td>6. Vocabulary service implemented</td>
<td></td>
</tr>
</tbody>
</table>

[Source: [Final Report](http://ands.org.au) | [Milestone Date](http://ands.org.au) | [2013-14 Milestones](http://ands.org.au) | [Progress](http://ands.org.au)]
3.5 ARDC Application Infrastructure

These are the milestones for the Institutional Engagement Program in the 2013-14 Annual Business Plan. Progress is reported against the NCRIS Seeding the Commons Program, the EIF Data Capture Program, the EIF Metadata Stores Program, and the ARDC Application Program.

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>13Q3</strong></td>
<td>Five institutional agreements on work programs for their research data assets</td>
<td>Institutional agreements on data assets were not worked on in ARDC Applications program</td>
</tr>
<tr>
<td></td>
<td>Successful conclusion of all Data Capture, Seeding the Commons, Metadata Stores and Applications projects</td>
<td>9 of the 25 Applications projects completed</td>
</tr>
<tr>
<td><strong>13Q4</strong></td>
<td>Further five institutional agreements on work programs for their research data assets Identification of some institutional assets</td>
<td>Institutional agreements on data assets were not worked on in ARDC Applications program</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20 Applications projects completed</td>
</tr>
<tr>
<td><strong>14Q1</strong></td>
<td>Further five institutional agreements on work programs for their research data assets Identification of some institutional assets Concluded agreed work on research data assets with initial five institutions</td>
<td>Institutional agreements on data assets were not worked on in ARDC Applications program</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20 Applications projects completed</td>
</tr>
<tr>
<td><strong>14Q2</strong></td>
<td>Further five institutional agreements on work programs for their research data assets. Identification of some institutional assets. Concluded agreed work on research data assets with further five institutions</td>
<td>Institutional agreements on data assets were not worked on in ARDC Applications program</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 25 Applications projects completed</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><strong>13Q3</strong></td>
<td>Second Research Data Alliance Plenary takes place Linguistics Working Group approved by Research Data Alliance Council</td>
<td>Plenary took place as planned Linguistics Working Group approved</td>
</tr>
<tr>
<td><strong>13Q4</strong></td>
<td>Another Working Group with significant Australian involvement approved by Research Data Alliance Council ANDS has contributed to a major ODIN event</td>
<td>Another Working Group not approved, but significant progress made towards group focussed on Data Description Registry Interoperability</td>
</tr>
<tr>
<td><strong>14Q1</strong></td>
<td>ANDS hosts the Third Plenary in Europe Third Research Data Alliance Plenary takes place</td>
<td>ANDS, together with Ireland, hosted the Third Plenary in Dublin</td>
</tr>
<tr>
<td><strong>14Q2</strong></td>
<td>First Research Data Alliance Working Groups deliver data exchange solutions ANDS has delivered a draft ODIN report</td>
<td>First Working Groups will now deliver in September 2014 ANDS contributed significantly towards the ODIN draft report, as well as to planning for its successor project</td>
</tr>
</tbody>
</table>
4. Deviations from the Project Plan

The ANDS Business Plan for 2013-14 had to take into account NCRIS, EIF, CRIS, DWF, and NCRIS 2013 funding, and this report describes activity against the NCRIS and EIF components of this funding. It has thus had to separate activity in an integrated set of activities based on the funding source, and being mindful of the differing conditions of grant.

There have been no substantial changes or deviations from the business plan that was submitted, but the focus on institutional engagement during the year moved from completion of funded institutional research data infrastructure projects, to broader institutional engagement, and the establishment of NCRIS 2013 funded major open data collections. These collections diluted to some extent focus on more general collection development to focus on open data collections. This became particularly important given the growing strength of open data expectations in Australia and internationally.

ANDS also crafted a considerable response to the ARC funding route changes that were not anticipated in the business plan but were consistent with the business plan.

[Previously] There was no system for centrally documenting and recording the existence of research datasets at the University [of Western Australia], or for publicising their existence nationally and internationally. Through the implementation of the VIVO metadata hub, the University now has a platform for recording and publicising datasets, and for contributing descriptions of datasets to Research Data Australia. A total of 83 dataset descriptions were made available through Research Data Australia as a result of this project.
5. Financial and Human Resources

The following table indicates ANDS expenditure by program for July 2013 to June 2014. Income is shown in the audit statement in Section 8.

<table>
<thead>
<tr>
<th>ANDS Expenditure for the Financial Year ended 30th June 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item</strong></td>
</tr>
<tr>
<td>SALARIES</td>
</tr>
<tr>
<td>Applications</td>
</tr>
<tr>
<td>Core</td>
</tr>
<tr>
<td>Data Capture</td>
</tr>
<tr>
<td>Institutional Engagement</td>
</tr>
<tr>
<td>Metadata</td>
</tr>
<tr>
<td>Project Office</td>
</tr>
<tr>
<td>Public Data</td>
</tr>
<tr>
<td><strong>Total Salaries</strong></td>
</tr>
<tr>
<td>OPERATING EXPENSES</td>
</tr>
<tr>
<td>Applications Internal Program Expenses</td>
</tr>
<tr>
<td>Applications Contracts</td>
</tr>
<tr>
<td>Core Contracts</td>
</tr>
<tr>
<td>Core Internal Program Expenses</td>
</tr>
<tr>
<td>Data Capture Internal Program Expenses</td>
</tr>
<tr>
<td>Data Capture Contracts</td>
</tr>
<tr>
<td>Institutional Engagement Internal Program Expenses</td>
</tr>
<tr>
<td>International Collaboration Internal Program Expenses</td>
</tr>
<tr>
<td>Metadata Internal Program Expenses</td>
</tr>
<tr>
<td>Metadata Contracts</td>
</tr>
<tr>
<td>Public Data Internal Program Expenses</td>
</tr>
<tr>
<td>Public Data Contracts</td>
</tr>
<tr>
<td>Public Data CSIRO Program Funding Payment</td>
</tr>
<tr>
<td>Project Office Internal Expenses</td>
</tr>
<tr>
<td>Project Office CSIRO Support Costs</td>
</tr>
<tr>
<td>Project Office ANU Support Costs</td>
</tr>
<tr>
<td><strong>Total Operating Expenses</strong></td>
</tr>
<tr>
<td><strong>EIF Total Expenditure</strong></td>
</tr>
</tbody>
</table>

*Table 4: ANDS Expenditure for 2013-14*
Staffing and recruitment has been extensive since the establishment of the project. The organisation chart at the end of June 2014 is shown in Figure 15.

Figure 15: ANDS Organisational Chart
6. Co-Investment

6.1 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.

6.2 Project Co-Investment

As a result of the original 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 was changed to a more flexible arrangement where ANDS contracted with partners to allow for a mix of staffing needs.

ANDS investments at institutions in Seeding the Commons and Data Capture activities have triggered substantial co-investment and post-investment with a number of ongoing or fixed term positions being created, and over $3.8M of institutional investment made to date, and over $3M of post-project investment identified to date. This indicates the extent to which institutions are embedding 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 had commitments of around $3.8 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 $75,431,120
- Total Partner co-investment $13,993,316
- Total Post project co-investment is estimated to be 18.55% of ANDS investment

Looking just at the financial year 2013-14:

- Total ANDS Funds $18,972,670
- Total Partner co-investment $3,367,866
- Total Post project co-investment is estimated to be 17.75% of ANDS investment
7. Performance Indicators

7.1 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:** 71 institutions fed collection descriptions to RDA along with 46 individual collections. From these institutions, 157 data source feeds have been setup (101 automatic and 56 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.
- Publicly Funded Research Organisations: ANSTO, CSIRO and AIMS.
- 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, 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 AustLit 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.

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.

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:** In November 2013, ANDS started using Google analytics for logging searches in Research Data Australia. From then on, there...
were a total of 23,769 searches in Research Data Australia, with an average of around 2,900 searches per month.

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: 310,697 page views and 200,760 unique page views (from Google Analytics tool). Unique page views increased 75% from last year, increased more than 4 times since June 2011.

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 101 respondents: including partners and researchers. Over 84% of the Universities that ANDS works with responded to the Survey.

### Key Outcomes:

| 1. Satisfaction with ANDS services overall | • 94% were satisfied or very satisfied  
|                                           | • 5% dissatisfied                      |
| 2. Awareness of the value of data        | • 81% 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 | • 75% noticed a change (up from 63% last year)  
|                                           | • 13% partner) noticed no change (down from 18% last year)  
|                                           | • 13% were unsure (down from 20% last year) |
| 4. Satisfied with Research Data Australia (RDA) | • 75% satisfied or very satisfied |
| 5. Research Data routinely deposited into data repositories | • 20% usually or always  
|                                         | • 58% seldom |

Compared to the last survey ANDS has kept the level of satisfaction stable at 82%, a slight rise (3%) from prior year.
ANDS helped to share contacts and put me in contact with others who might have similar experiences. For example, TERN developed a functioning DataCite API program. ANDS suggested I make contact with TERN to discuss further rather than writing our own new utility.

DR JINGBO WANG, DATA COLLECTIONS MANAGER, NATIONAL COMPUTATIONAL INFRASTRUCTURE (NCI)

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 71 organisations to publish data collection descriptions in either the public or draft systems, 3 new agreements added since 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: 97,672 collections are made available in Research Data Australia as at 30 June 2014. 89% of the total collection records came from the top 5 top contributors: 22% from Geoscience Australia (GA), 19% from Queensland Facility for Advanced Bioinformatics (QFAB), 17% from State Records NSW, 16% from Public Records Office Victoria (PROV) and 15% from Australian Ocean Data Network (AODN). The remaining 11% of the total collections are from 66 contributors with a total of 85 corresponding groups appearing in Research Data Australia.

The number of ANDS research data sets increased 45% from June last year.

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

Result: 12,082 persistent identifiers (9,677 PIDS handles, 2,405 DOI handles), 17% higher than June last year. These identifiers were minted by 17 Identify My Data (PIDS) clients and 13 out of 29 registered 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 also help inform future surveys.
7.2 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 2013-14, 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>13-14</th>
<th>Maximum 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Institutions Participation in RDA ¹</td>
<td>NA ²</td>
<td>21</td>
<td>35</td>
<td>40</td>
<td>41</td>
<td>43</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>25</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>71</td>
<td>43</td>
</tr>
<tr>
<td>Institutional Research Metadata Store</td>
<td>0</td>
<td>9</td>
<td>25</td>
<td>25</td>
<td>31</td>
<td>43</td>
</tr>
<tr>
<td>Research Data Provider Participation ²</td>
<td>NA ³</td>
<td>9</td>
<td>30</td>
<td>28</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td>Research Data Infrastructure Partners ³</td>
<td>4</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>26</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>97,672</td>
<td></td>
</tr>
<tr>
<td>Research Data Exploitation Tools</td>
<td>9</td>
<td>26</td>
<td>30</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Fields of Research Coverage 4</td>
<td>5</td>
<td>21</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 5: 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
In previous Annual Reports we were able to report on some overarching achievements:

- A populated Australian Research Data Commons (ARDC) has been established.
- 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.
- Data is overwhelmingly on the agenda in research and research infrastructure, and ANDS has helped position Australia internationally.
- The Australian Research Data Commons has matured and grown substantially.

There were a number of notable changes in the research data environment that are worth noting as they had an influence on ANDS activities:

- ARC change in Policy to require data planning as part of the grant submission process.
- The European Horizon 2020 funding program allocating £1.4b to open data.
- Publishers increased involvement with research data.
- The increased level of commercial data services such as Figshare and Zenodo.
- The advent of RDSI data storage availability - at a level and scale that was very different a year ago.
- Substantial acceptance of the need for data licencing such as through AusGOAL.

The major activities for ANDS as a whole in the 2013-14 calendar year were:

- Making more than 97,000 research data collection pages discoverable through Research Data Australia, Google, and other search engines harvested by ANDS at over 70 research data providing institutions.
- Adoption of the data citation services by 25 research institutions.
- Maturing research data management community strongly engaged with ANDS with substantial increase participation in, and demand for, ANDS events.
- Greater contribution to the national policy environment as exemplified by contributions to the Chief Scientist’s Office, National Environment Research Program, Research Data Infrastructure Committee.
- 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.
- Completed a program of installation of research data infrastructure at institutions through the Seeding the Commons, Data Capture, Metadata Stores, and Data Applications programs.
- Building the RD-A and subsequent growth (Plenaries: 230, 365, 500, 700) 8 founding members to 300 to now over 1000 member organisations.
- Engagement with the research institutions in general and the research offices in particular around the rule changes by the ARC.

Research Data Australia is not only a discovery portal, but also a publicity portal. It’s a way to make research data visible to others.

DR ROSS WILKINSON
8. 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 2014

<table>
<thead>
<tr>
<th>Item</th>
<th>2013</th>
<th>Appendix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Grant</td>
<td>-</td>
<td>A</td>
</tr>
<tr>
<td>Commercial</td>
<td>192</td>
<td>A</td>
</tr>
<tr>
<td>Interest</td>
<td>156,943</td>
<td>A</td>
</tr>
<tr>
<td>Total Income (a)</td>
<td>159,535</td>
<td>B</td>
</tr>
<tr>
<td>Expenses</td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>Salaries</td>
<td>1,304,675</td>
<td>B</td>
</tr>
<tr>
<td>Non-Salary</td>
<td>8,713,010</td>
<td>B</td>
</tr>
<tr>
<td>Total Expenses (b)</td>
<td>10,107,685</td>
<td>C</td>
</tr>
</tbody>
</table>

Surplus/(Deficit) for the reporting period (a - b)  
Brought forward surplus/(deficit) from 30 June 2013 (c)  
Balance Carried Forward to next Reporting Period (a - b) + (c)

10,948,150  
9,946,150  
0

We, Ian Smith 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 period ended 30 June 2014.

(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 IAN SMITH  
Vice-Provost, Research & Research Infrastructure  
Office of the Provost and Senior Vice-President, Monash University

MR JOEL CHIBERT  
Director, Research & Revenue Accounting Services  
Office of the Chief Financial Officer and Senior Vice-President, Monash University
9. Summary of the Conduct of the Project

The ANDS project commenced on January 1st 2009, as a result of an extensive consultation that lead to the “Towards the Australian Data Commons Report” that set the directions of the project, and an extensive process in the ANDS establishment process that determined governance, management and processes for the project, and determined the ANDS project plan. The project commenced with offices in Melbourne and Canberra, and staff from Monash, ANU, ad CSIRO, working through four programs:

• Seeding the Commons.
• Developing Frameworks.
• Building Capabilities.
• Providing Utilities.

The first major engagement was a consultation around the country which lead to an important decision: the long term custodians of research data are largely long lived institutions, so ANDS should focus its efforts on support of research institutions, and research data providing institutions, and support institutions in their direct engagement with researchers.

By June 30th 2009, partnerships were being established with research institutions, local and regional eResearch providers, research data providers, and NCRIS partners, particularly ARCS as we jointly ran the NeAT projects. By that date:

• An operational discovery service has been created.
• Initial research collections from IMOS, TARDIS and iVEC were discoverable.
• Initial material was developed to support research institutions developing their own research data capability.
• Guidelines on how to comply with the data management requirements in the Australian Code were prepared.

• NeAT projects were established and managed.

As a result of the Super Science program being announced in May 2009, the work of ANDS needed to rapidly change in size ($24M to $72M); in nature (from an organization that worked more directly on the changes to Australia’s research data environment to one that funded partner organisations to do the bulk of the work); and in time (from two years to four).

The major activities for ANDS as a whole in 2009-10 were:

• Establishing relationships with all of our key partners in making the ARDC a reality.
• Establishing processes to fund programs effectively and efficiently.
• Initiating a $10M set of activities in ARDC called Fast Start activities that enabled early progress on a 2 year program.
• Establishing infrastructure to identify, register and publish collections descriptions through Research Data Australia, the ANDS portal into the Australian Research Data Commons.
• Commencing a $16.15M EOI process for both Data Capture and Seeding the Commons based on research output, with appropriate checks and balances to enable good outcomes and accountability.
• Continuing the NeAT program of discipline enhancing tools for improved collaboration and exploitation of research data.
• Establishing a set of landmark public sector data activities to make public data available to researchers.

Whilst there was inevitable disruption with the changes described and this has slowed some delivery, it has been very helpful in deepening engagement with our partners, and increasing the scope of data capture and sharing that had previously been envisaged in the Seeding the Commons program.

The breakthrough that occurred in 2009-10 was that research data management and ANDS has moved from a “good idea” but with little institutional engagement, to a situation where 33 Universities and many other organisations are actively engaged on data transformation projects at their institutions, and where institutional investment in managing their research data is occurring at the policy, procedural and practical infrastructure. ANDS has been a catalyst for much of this activity and research data collections are recognized now as key to many institutions’ research ambitions.

By June 30th, 2011, the Australian Research Data Commons (ARDC) had been established. 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 re-use, 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. This included:

• Enabling research data collections to be described and harvested by ANDS at 26 research institutions.
• Enabling research data collections to be described and harvested by ANDS at 27 research data providers including other research infrastructure providers, public sector providers, cultural institutions and research consortia.
Making 26,746 research data collection from 21 providers discoverable.

By June 30th 2012, there had been a greatly increased research data management capacity at our research institutions. ANDS estimates that there were approximately 300 people working on data management within research institutions, which is probably a tenfold increase compared to January 2009. Research institutions are seeing substantial value in this infrastructure. 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 were embedding a research data infrastructure into standard operations.

Significant international activity took place focused on interacting with the European Commission, national activity and EC funded projects. Through engagement initiated by the Department leading to a meeting on research infrastructure, ANDS participated in discussions leading to establishing Research Data Alliance, where Australia is partnering with the US and the EU.

The major activities for ANDS conducted by June 30th, 2013 were:

- Made more than 70,000 research data collections discoverable through Research Data Australia.
- All major research institutions had installed substantial research data infrastructure, notably metadata stores, that enabled them to participate in the Australian Research Data Commons.
- Demonstrated the value of reusing data with a suite of applications that enabled high profile researchers to provide compelling demonstrations of new ways of conducting data intensive research.
- Established 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 attracted a very substantial level of international commitment.

Taken together, two important changes occurred during the final year of the project – firstly there was 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 focus of thinking about research data and having conversations, developing policy responses, and involving all relevant agencies, though ANDS policy and practice capabilities.

By June 30th, 2014 – the final year of the project, all components of the project had completed successfully. A summary of the year’s activity is described in the previous chapter. Perhaps most notably, the Australian Research Council made a major change to its funding rules by requiring statements about research data management to be made at the time of funding submissions, and the research sector was ready. ANDS immediately ran a comprehensive engagement supporting research institutions in their engagement with their researchers. ANDS had been engaging at many levels on research data policy, enabling a substantial change to the research data landscape in Australia.

In summary, Australia has a coherent approach to research data, has an international profile enabling Australia’s researchers and research institutions to form partnerships internationally to tackle the most demanding data intensive research. Much needs to be done – while this advantage has been established, the data collections themselves had not been established to the extent necessary or desirable, and partnership advantages have not been exploited to the extent possible. Nevertheless, Australia has established an international competitive advantage in research data.

I know that our society and our future are best served through science that is faster, better and more open.

NEELIE KROES, VICE PRESIDENT OF THE EUROPEAN COMMISSION
10. Expected Future Usage of the Infrastructure

Research data, and research data infrastructure will be vital for the best research to be conducted, so will continue to be used. Increasingly there is an expectation that publicly funded research will make its outputs, including research data publicly available at no cost. The value to the nation of its research data assets being available publicly are very high, as is the value to research institutions. It is thus highly desirable that research data infrastructure continues to be available at no cost, to ensure that researchers do not see barriers to publicly sharing data generated through their projects. Equally important, the need for freely accessible means of publishing research data, enabling researchers to reap the citation benefits, builds the value of open data to researchers.

[Semaphore] really makes my work a lot easier, also just to enable us to store our data and our modeling results in a very efficient way that is accessible for researchers in Australia, and actually worldwide.

DR CLEMENS SCHEER, SENIOR RESEARCHER
QUEENSLAND UNIVERSITY OF TECHNOLOGY, ON SEMAPHORE
Internationally the momentum to conduct data intensive research is increasing, and international effort to establish improved access to data, through open data policy, and improved data infrastructure, is also increasing. The rapid growth of the Research Data Alliance, and the G8 Science Ministers’ declarations are strong evidence of these trends. As well, within domains, the staggering growth of publicly available data through such initiatives as EMBL and PDB, or data being generated by CERN, demonstrate just how important it is to be a leader in data intensive research. The development of the SKA with its staggering levels of data is both and Australian opportunity and a challenge. Many areas of research that used to have relatively little digital information are being transformed by new data capture approaches. Archaeology is using three-dimensional laser scanning of megalithic rock art, LiDAR data, and underwater video to capture data, leading to substantial infrastructure challenges.

Australia will see the same challenges, and will need to assemble the data resources to tackle its own biggest challenges, seeking the best international partners, with infrastructure that enables partnership over data that cannot move due to its size.

Research data must become the means that enables research, industry, education, government and public to engage – over policy, over evidence, to support training, to enable innovation, and to solve the national challenges. Research data must thus be available with rich context beyond the purposes for which it was collected.

Annually, Australia may spend up to $1B on new research data – it must get the best value from that data – initially with the best researchers partnering with a data toolset, and beyond as data is assembled for new purposes.
12. Appendices

12.1 Confidential Information

There is no confidential information.

12.2 Progress against activities

12.2.1 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>
</tbody>
</table>

The remaining funds in the Data Capture program will be allocated according to two criteria:
- 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.
- 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.

Excess funding was redirected to other ANDS programs, in consultation with the Steering Committee.

12.2.2 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>
</tbody>
</table>
**Proposed Activity**

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.

**Progress**

ANDS opted not to commission development in this area in order to release funds for institutional collection metadata stores.

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.

**Progress**

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.

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.

**Progress**

ANDS support has taken the form of funding to 22 institutions to improve their existing metadata store infrastructure. Work with RDSI nodes has transferred to the National Collections program, although some support is provided by client liaison officers across ANDS.

### 12.2.3 Public Sector Data Access Infrastructure

<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 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.</td>
</tr>
<tr>
<td>Agency or Institution or Project</td>
<td>Status and Description</td>
<td>Progress</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<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>
<td>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 has 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.</td>
</tr>
<tr>
<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 made available.</td>
<td>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.</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 back sets. This data will also be contributed to Research Data Australia and the ARDC and when finalised will provide a comprehensive longitudinal view.</td>
<td>Completed</td>
</tr>
<tr>
<td>Agency or Institution or Project</td>
<td>Status and Description</td>
<td>Progress</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Museums Metadata Exchange</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>National Archives of Australia</td>
<td>Proposal reviewed and value of commitment revised. NAA are re-assessing their intentions.</td>
<td>See PROV below</td>
</tr>
<tr>
<td>Public Records Office of Victoria</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. 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.</td>
</tr>
<tr>
<td>GeoSciences Australia</td>
<td>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. 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.</td>
<td></td>
</tr>
</tbody>
</table>
12.2.4 ARDC Core Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Support complex national collections</td>
<td>• Annotations implemented</td>
</tr>
<tr>
<td>• Annotations for distributed collections</td>
<td>• Duplicate record strategy implemented</td>
</tr>
<tr>
<td>• Display of duplicate records</td>
<td>• Hierarchical viewer implemented</td>
</tr>
<tr>
<td>• Collection structural hierarchy and connections viewer</td>
<td>• Crosswalks implemented</td>
</tr>
<tr>
<td>• Support crosswalks from community schema to make collection registration easier</td>
<td>• ARC and NHMRC grant information linkable</td>
</tr>
<tr>
<td>• Establish processes for all ARC/NHMRC grant information to be easily linkable to data collections information</td>
<td>• Publications linked</td>
</tr>
<tr>
<td>• Test processes for journal publications information to be easily linkable to data collections information</td>
<td>• SEO optimised</td>
</tr>
<tr>
<td>• Optimise search engine indexing</td>
<td>• Harvester service re engineered</td>
</tr>
<tr>
<td>• Continue to support and operate the harvesting and registration service</td>
<td>• Discovery portal maintained</td>
</tr>
<tr>
<td>• Continue to operate the discovery portal</td>
<td>• Data Citation and publication system materials provided</td>
</tr>
<tr>
<td>• Work with publishers, data archives, and citation indexes to establish a data citation as a routine scholarly practice</td>
<td>• DOI service operational</td>
</tr>
<tr>
<td>• Optimise self service and automated processes for DOI updating (persistent identification)</td>
<td>• Extensive work with publishers</td>
</tr>
<tr>
<td>• Test integration of registry and DOI services (simultaneous mint and register)</td>
<td>• DOI service optimised</td>
</tr>
<tr>
<td>• Establish national vocabulary catalogue</td>
<td>• DOI integration investigated</td>
</tr>
<tr>
<td>• Establish national vocabulary service</td>
<td>• Beta Vocabulary Catalog established</td>
</tr>
<tr>
<td>• Launch, procedures, promotion and support</td>
<td>• Production Vocabulary Service established</td>
</tr>
<tr>
<td>• Integrate ANDS systems with ORCID</td>
<td>• Procedures promotion support project launched</td>
</tr>
<tr>
<td>• Support the use of researcher identifiers and good practice encoding</td>
<td>• Integration with ORCID complete</td>
</tr>
<tr>
<td>• Test linkages with Google Scholar</td>
<td>• Research identifiers supported through ODIN project and ORCID support</td>
</tr>
<tr>
<td>• Component plug and play architecture</td>
<td>• Google Scholar linkages tested</td>
</tr>
<tr>
<td>• Component architecture phase one complete</td>
<td>• ANDS Services help desk</td>
</tr>
<tr>
<td>• Community support and communications established</td>
<td>• ANDS Technical guides</td>
</tr>
<tr>
<td>• Support partners providing information about their data assets to the national registry and Research Data Australia</td>
<td>• ANDS Developers Toolbox</td>
</tr>
<tr>
<td>• Support partners implementing data publication and promoting data citation through the use of the ANDS DOI Service</td>
<td></td>
</tr>
<tr>
<td>• Support the establishment of the ANDS vocabulary service and support for the initial adopters group</td>
<td></td>
</tr>
</tbody>
</table>

12.2.5 ARDC Applications Infrastructure

As noted in the Business Plan for 2013-14, the Applications program was concluding this year, though there was a small amount of remaining activity being managed in the Institutional Engagement Program. As such, there was not a separate Activity section for Applications in the Business Plan. There was a single, effectively continued from the previous year, activity, which was this:

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrate the value of better use of research data through the Application program and associated activities</td>
<td>For evidence of the successful completion of this activity, please consult the descriptions of the applications projects under Institutional Engagement.</td>
</tr>
</tbody>
</table>
### 12.2.6 International Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work to establish and develop the Research Data Alliance (RD-A)</td>
<td>Research Data Alliance established and growing very healthily!</td>
</tr>
</tbody>
</table>
| Contribute to ongoing operations of Research Data Alliance | Contributed to RD-A Organising Group meetings and associated activities  
Contributed to the RD-A Secretariat  
Took part in Council Meetings  
Contributed to other supporting bodies of the RD-A, as required  
Encouraged Australians to join and contribute to Interest and Working Groups, as required. Also worked to establish WG around Data Registration Interoperability  
Developed and refined policy and operations documents  
Liaised with specific Working Groups and Interest Groups to assist them to engage with RD-A processes  
Took responsibility for the organisation of the Third Plenary (by chairing the organising committee) which was held in March 2014  
Contributed to the process of setting up a legal entity for RD-A  
Facilitated involvement of other countries and institutions in RD-A |
| Facilitate Australian involvement in Research Data Alliance working groups | Worked with Linguistics community to develop proposal for ISOCat language codes as RD-A working group, and nurture this group through first year  
Worked with Marine community through IMOS as they developed a proposal for a Marine RD-A working group  
Actively sought and worked with other Australian representatives of other communities for whom an RD-A working group is an appropriate vehicle to develop and implement international data exchange technologies – in particular a group around Urban Quality of Life Indicators |
| Contribute to success and expansion of persistent identification of data through DataCite | Attended DataCite Board Meetings and Annual General Meetings  
Promoted DataCite as a technology for data citation (see other references elsewhere in the report for specific activity in this area) |
| Contribute to success of data and research identification integration through ODIN | Delivered against work packages assigned to ANDS in ODIN project plan; worked to ensure successful completion of ODIN |
| Align work plans between British and Dutch organisations, including the JISC-MRD program, DCC, and DANS | Shared ANDS work plans (including this document) with Digital Curation Centre (UK), Data Archiving and Network Services (NL), JISC Managing Research Data program (UK) and SURF (NL) either no longer in existence or have significantly wound back their data involvement |

Data storage is of particular concern for big data generators, such as these five antennae of the Australia Telescope Compact Array, near Narrabri, NSW. Image courtesy of CSIRO and John Masterson (CC BY 3.0)
12.3 Risk Register

ANDS maintains a Risk Register. The risk assessment methodology, adapted from the Australian Risk Management Standard AS/NZS 31000:2009, 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.

12.3.1 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.

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.
12.3.2 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.

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.

12.3.3 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).
12.3.4 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.

Risk 12 – That ANDS is most effectively reporting under all the various programs

Risk Factors:
- Staff needing to report on similar activities multiple times throughout the year as per funding agreements that have been signed with the Commonwealth Government.

Risk Mitigations:
- Develop a reporting calendar and circulate amongst relevant staff and provide reminders.