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

Annual Report 3, 30 September 2012
TABLE OF CONTENTS

1 PROJECT STATUS ............................................................................................................ 4
  1.1 Background ................................................................................................................. 4
  1.2 Major Activities, Breakthroughs, Highlights, and Issues ............................................. 6

2 ACTIVITIES UNDERTAKEN ............................................................................................ 11
  2.1 Research Infrastructure ............................................................................................. 11
  2.2 Data Capture ............................................................................................................... 12
  2.3 Metadata Stores ......................................................................................................... 26
  2.4 Public Sector Data ..................................................................................................... 30
  2.5 ARDC Core Infrastructure .......................................................................................... 35
  2.6 ARDC Applications ................................................................................................... 50
  2.7 Project Office ............................................................................................................ 53
  2.8 Promotion .................................................................................................................. 54
  2.9 Risk Management ...................................................................................................... 60

3 PROGRESS AGAINST MILESTONES ............................................................................ 61
  3.1 Data Capture Infrastructure ....................................................................................... 61
  3.2 Metadata Store Infrastructure .................................................................................... 62
  3.3 Public Sector Data Infrastructure ............................................................................... 63
  3.4 ARDC Core Infrastructure ......................................................................................... 64
  3.5 ARDC Application Infrastructure ............................................................................... 65
4 DEVIATIONS FROM THE PROJECT PLAN .................................................................67
5 COMMITMENTS ..................................................................................................70
6 FINANCIAL AND HUMAN RESOURCES ..............................................................71
7 CO-INVESTMENT .................................................................................................74
  7.1 Access and Pricing .........................................................................................74
  7.2 Project Co-Investment ....................................................................................74
8 PERFORMANCE INDICATORS .............................................................................75
  8.1 KPI Report ....................................................................................................75
  8.2 Progress over the Life of the ANDS Project ....................................................78
  8.3 Overall Progress ............................................................................................79
9 AUDIT STATEMENT ............................................................................................80
10 APPENDICES .....................................................................................................81
  10.1 Confidential Information ..............................................................................81
  10.2 Project Description Detail ............................................................................81
  10.3 Progress against activities ..........................................................................112
  10.4 Risk Register ...............................................................................................120
1 Project Status

1.1 Background

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

The Australian National Data Service 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 is continuing to increase. In May 2010 the Australian Research Data Commons (ARDC) project was announced as an EIF funded Super Science Initiative and 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. Most recently, ANDS extended its operations beyond June 2011 to June 2013. Each of these changes has had a high impact on the activities of ANDS and the second change has substantially affected the ANDS project as agreed in the 2011-12 Business Plan. As each of these changes has occurred, ANDS has continued to manage the ANDS and ARDC projects together, as they are strongly co-dependent. There has been no significant variation subsequent to this extension, so this report describes activity taking place against the agreed 2011-12 Business Plan.

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 consolidated into two programs of activity:

- **Frameworks and Capabilities** – the frameworks that will enable research data producing institutions to capture, manage and share research data; and improving Australia’s capability to manage its research data
- **Seeding the Commons** – improving local data capture and populating the data commons

The associated ARDC project has five programs of activity:

- **Data Capture** – an institutionally based program to automate the capture of data and metadata from instruments (broadly defined) in data intensive research
Public Sector Data – a program of making more public data collections visible and available through the ARDC

Metadata Stores – an institutionally based program that enables metadata to be stored coherently across an institution that supports data management, publishing, sharing and reuse

ARDC Core Infrastructure – an ANDS driven program that puts in place the national services that enable research data to be published and discovered (It is an expansion of the Providing Utilities program)

ARDC Applications – a program that develops tools and services to support demonstrations of the value of exploiting data in the ARDC

Two new programs were subsequently created in 2012:

National Collections - an ANDS-driven, NCRIS-funded program partnering with institutions wishing to make National Collections available, and with RDSI and its nodes to help improve storage and access to those collection.

International Infrastructure - a program designed to work collaboratively with international organisations and partners to ensure a more compatible international data-sharing environment for Australian researchers

These programs were created to respond to changing environment, rather than changing focus of ANDS. The advent of the RDSI initiative meant that there was new opportunity to assemble and make available collection 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 that are managed connected discoverable and increasingly usable through better access, and possibly with new forms of access.

The strong drive in both Europe and the US in particular but 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. It 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 (TADC), 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.” – TADC

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
The ARDC investment statement can be seen as an intensification of effort in support of the second TADC statement. This report describes progress against this intent.

1.2 Major Activities, Breakthroughs, Highlights, and Issues

In the previous Annual Report we were able to report two headline overarching achievements:

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

“Working with ANDS has brought the issues associated with research data access and management home to the university. ANDS has opened up new research pathways that will give us better outcomes from our large data sets both locally and nationally. This is particularly so in areas that such as environmental monitoring and plant breeding data associated with the Plant Accelerator. Working with ANDS has kicked off a culture change within the University which will have a tremendously positive impact on the way we value and utilise research data as a resource for the future” – Professor Mike Brooks, DVCR, University of Adelaide
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.

By June 2012, three further key changes have occurred in Australia’s research data environment:

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

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

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

The major activities for ANDS as a whole in the 2011-12 calendar year were:

With no further project effort, there has been a three-fold increase in collections available from CSIRO in Research Data Australia as a collection become available in the CSIRO Data Access Portal.

"...the main challenges therefore relate to data... The definite “good news” story for Australia is that ANDS, which is an established project, recognises this problem and is clearly recognised as a leader in finding the solutions by similar programs overseas. ANDS, in combination with a newer program NeCTAR, supported by hardware infrastructure through NCI and RDSI genuinely seem ahead of their equivalents overseas" - Professor Andy Pitman, UNSW

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

The major activities for ANDS as a whole in the 2011-12 calendar year were:
Making 40,811 research data collection pagesdiscoverable through Research Data Australia, Google, and other search engines harvested by ANDS at 62 research data providing institutions.

Augmentingestablishing infrastructure to identify, register and publish collections descriptions through Research Data Australia with the ANDS Cite my Data service.

Furtheringacoherent approach to research data management at all major research institutions, along with the tools and technologies that enable them to participate in the Australian Research Data Commons, particularly through the Metadata Stores being installed at the 25 most research intensive institutions.

Building asuite of applications that demonstrates the value of reusing data that is combined from many sources to enable high profile researchers to explore new problems in new ways.

There have been many highlights over the past year. They can be summarised as the establishment of coherent institutional research data infrastructure, making research data a national research advantage, and working to bring about a mature Australian Research Data Commons.

Establishing coherent institutional research data infrastructure:

- Institutions are increasingly taking an overarching responsibility for their research data, rather than seeing it as a researcher responsibility, meaning that institution-wide approaches are being adopted.
- More tools have been deployed to automatically capture rich metadata along with the data for a wide range of instruments
- More institutions are deploying metadata stores, connected to institutional and research systems
- More institutions have operational research data management plans
- More institutions are appointing staff with an on-going data management role
- Institutions are increasingly collaborating on approaches and learning from each other – there is rarely a unique solution being used at an institution
- 1,098 people have attended ANDS organised events in January-September 2012

Helping to establish research data as a national research advantage:

- Positioning Australia as a key participant in the newly forming Research Data Alliance

30 September 2012

ands.org.au
- Research data features prominently in both the National Research Infrastructure Investment Plan and the 2011 Research Infrastructure Investment Plan
- ANDS has worked actively in partnership with the key data generating infrastructure facilities, such as AURIN, TERN and IMOS
- ANDS has been working with others to build a policy framework to enable much simpler access and reuse of research data
- ANDS material, especially the commissioned report on the economic benefit of free access to data, is being used and cited internationally

A mature Australian Research Data Commons:

- 40,811 collections were described and available through Research Data Australia by June 30th, 2012
- Over 62 research data providing institutions
- Data from all Fields of Research is discoverable
- Enhanced publication services and discovery services, with the addition of the Cite my Data Service
- Improved connectivity with the new Gazetteer service enabling easier discovery and linking of data by location
- Enhanced capacity to use the data commons with better data, more sophisticated tools and a range of Applications projects focused on demonstrating the value of this capacity

The 2011-12 business year had a large number of activities, but saw no major change in direction or approach. ANDS needed to respond to a developing international engagement opportunity and an opportunity to increasing focus on collections of significance beyond a single institutional as a result of the RDSI opportunity with the possibility of a richer toolset being realised through the Nectar initiative. During this substantial set of activities ANDS confronted a number of issues that it had to deal with and learn from:

- The major challenge facing ANDS this year has been to work with a large number of partners delivering a large number of data infrastructure components simultaneously around the country
- ANDS has also needed to transition beyond project support to institutional partnerships in developing coherent approaches, whilst supporting the rapid start of many metadata storage solutions and the advent of new data storage options through RDSI.
- The ANDS approach of focusing on institutional partnerships has contrasted with other approaches, but rather than being challenged by them has been increasingly valuable as a complement to researcher or disciplinary foci.

To now summarise the progress of the establishment of the ARDC, Figure 2 shows the many components of the ARDC that have been completed by 30 June 2012 or are currently under development at that time (the
number in brackets). This emphasises just how much work is being conducted at and by our institutional partners and their e-research providers.

**Figure 2: Australia Research Data Commons progress**
2 Activities Undertaken

2.1 Research Infrastructure

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

Complementary infrastructure established as part of the NCRIS ANDS project is described in a separate progress report. Some of this infrastructure was to be established in the 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 infrastructure is being constructed through 5 programs:

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

A new program focused on **International Infrastructure** was introduced during the year to consolidate international activities and reflect an increasing focus on International Infrastructure.

The next section describes the specific research infrastructure created in the 2011-12 financial year.
2.2 Data Capture

2.2.1 Overview of program

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

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

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

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

2.2.2 Outline of projects

The intent of the Fast Start activities (further details for these can be found at Section 10.2.1) was twofold: to start expending the allocated funds (which at the time had to be expended by the end of 2010-11), thus smoothing somewhat the expenditure curve; and to quickly undertake a range of activities from which ANDS could learn and thereby fine-tune the process of expending the remainder of the Super Science funding. Given the short timeframe, ANDS sought and received approval from DIISRTE to select a small number of institutions where discussions were already underway and relationships had already been established.

For Data Capture, this involved the following institutions:
Details of these projects were all agreed by late 2010, and they are incorporated into the project table below.

The following NeAT projects were completed under the Data Capture program in 2010-11:

- Australian Node of the Human Variome Project
- Auscover Workflow
- Aus-e-Stage
- DataMiNX
- PODD
- Remote Tomography

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. Of the seventy projects agreed, 33 had been completed by June 30, 2012. A breakdown of the progress made in relation to this is provided below. A description of the projects underway in the reporting period is in section 10.2.3.

<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>X</td>
<td>X</td>
<td></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></td>
</tr>
<tr>
<td></td>
<td>Humanities and allied disciplines</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Project Title</td>
<td>Project agreed</td>
<td>Contracted</td>
<td>Underway</td>
<td>Completed</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>ANSTO</td>
<td>Metadata management for neutron beam instrument data - a joint project with Australian Synchrotron</td>
<td>X</td>
<td>X</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>X</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>X</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>X</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>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crystal Orientation Data Collection for Conversion to a General Data Type</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Flinders University</td>
<td>Automated measurement of the responses of wildlife populations to climate change</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Griffith University</td>
<td>Smart Water</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Adult Stem Cell &amp; Neurobiological Microscopy Instrumentation and Research Data Management</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>James Cook University</td>
<td>Tropical Data Hub</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>La Trobe University</td>
<td>CMSS RLI Metadata Capture and Publication</td>
<td>X</td>
<td>X</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>Papyri 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>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Biomedical Data Platform (Molecular Biology)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>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></td>
<td>Tools for curating and publishing research data in the form of media collections (Multimedia Collections &amp; ARROW)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Capture and publication of Australian ecosystem data from a network of measurement sites (Ecosystem Measurements)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Capture and publication of data on the history of adoption (History of Adoption)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Data Publication to Interferome (MIMR/Interferome)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Comprehensive Data Management for Microscopy Research Datasets</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Queensland University of Technology</td>
<td>Greenhouse Gas Emissions from Australian Soils</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Biodiversity</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>B150 Big Jam</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RMIT University</td>
<td>Data Capture from High Performance Computing Multi-User Environments</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Adelaide</td>
<td>Genomics Data Capture</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Automated capture and publishing of data generated on high throughput plant phenomic platforms.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Melbourne</td>
<td>Melbourne Neuropsychiatry Centre (MNC) Bioinformatics Development Project</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Youth Research Centre's Life Patterns Project: Longitudinal qualitative and quantitative survey data capture and reuse</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Video data in the Social Sciences. Optimising Metadata Capture, Data 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>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 New South Wales</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></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>X</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></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></td>
</tr>
<tr>
<td>University of Queensland</td>
<td>Spatially Integrated Social Science</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Microscopy/Microanalysis Image and Data Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DIMER Diffraction Image Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aquatic Species Tracking Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3D Anthropological and Archeological Collection Repository</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></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></td>
</tr>
<tr>
<td>University of Sydney</td>
<td>SKAMP Data Capture: astronomy</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NSW TARDIS Node</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>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 Tasmania</td>
<td>AgDataCapt: Capturing Agricultural Data</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>AMMRF Live Cell Microscope Data Capture</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Metadata Store/Aggregator</td>
<td></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></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Data Capture of state-wide hydrological datasets</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Technology, Sydney</td>
<td>Maximising the Benefit from Data-Intensive Processes at UTS</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Western Australia</td>
<td>Deployment and configuration of Institutional Metadata Repository</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Integrated Data Capture for Characterization and Analysis</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Archaeological Rock Art Data Capture</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Marine Ecology Video Capture and Storage</td>
<td></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></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Wollongong</td>
<td>Biomechanics Data Capture Project System</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Remote Sensing Spectral Library</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

**Table 1: Status of Data Capture projects**

### 2.2.3 Activity/Deliverables for 2011-12

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/](http://projects.ands.org.au/). Where no institution is named the project was completed as part of the NeAT program. As noted in Table 1, there were 70 projects underway or completed during the reporting period.
<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata management for neutron beam instrument data (ANSTO)</td>
<td>The challenges met by the MeCAT project included:</td>
</tr>
<tr>
<td></td>
<td>• choose a technology base to use for the project, either an existing open-source product or start from scratch. The choice was made to use the MyTARDIS (<a href="http://tardis.edu.au">http://tardis.edu.au</a>) software initially developed by Steve Androulakis at Monash University.</td>
</tr>
<tr>
<td></td>
<td>• help automate management of the data and metadata for the facilities</td>
</tr>
<tr>
<td></td>
<td>▪ for the Small and Wide angle X-ray scattering (SAXS/WAXS) and Infra-red Microspectroscopy (IR) beamlines at AS</td>
</tr>
<tr>
<td></td>
<td>▪ and the Echidna, Kowari, Platypus, Quokka and Wombat instruments at ANSTO;</td>
</tr>
<tr>
<td></td>
<td>• provide convenient remote access to the data for the users of the facilities</td>
</tr>
<tr>
<td></td>
<td>• improve metadata capturing, management and search</td>
</tr>
<tr>
<td></td>
<td>• facilitate publication of the data and registration in Research Data Australia</td>
</tr>
<tr>
<td></td>
<td>• create a community of MyTARDIS users to help maintain and extend the software</td>
</tr>
<tr>
<td>Metadata Capture and Storage for the Three Mature Beamlines at the Australian Synchrotron</td>
<td>See previous entry (combined project)</td>
</tr>
<tr>
<td>Aquatic Species Tracking Repository (University of Queensland)</td>
<td>This project is important to the conservation and management of native and feral species because it has delivered a platform that increases the understanding of animal behaviour. More specifically, the project has delivered:</td>
</tr>
<tr>
<td></td>
<td>• A platform for capturing and managing acoustic array and satellite data which includes: animal location data (lat, long, depth) as well as biological data (body temp) and environmental data (water temperatures, river heights/tidal flows and salinities);</td>
</tr>
<tr>
<td></td>
<td>• Open source software to enable the capture of animal movement data from multiple receiver arrays and its search, browse, retrieval and correlation with environmental, hydrological and oceanographic data;</td>
</tr>
<tr>
<td></td>
<td>• The development of the ECO-Lab’s data repository where metadata support fine-grained, spatio-temporal (GIS/timeline) discovery, decision-support and reuse, as well as the ability to filter, compress and control the quality of data sets that are</td>
</tr>
<tr>
<td>Project</td>
<td>Project Outcomes</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Greenhouse Gas Emissions from Australian Soils (QUT)</td>
<td>- The key achievements of this project are:</td>
</tr>
<tr>
<td></td>
<td>- The development of software to allow researchers to enhance metadata and manage the data automatically generated from instruments within the ecological research domain.</td>
</tr>
<tr>
<td></td>
<td>- Development of system architecture and suitable workflows for publishing research data information to Research Data Australia.</td>
</tr>
<tr>
<td></td>
<td>- A selection of 11 records described and contributed to Research Data Australia providing greater exposure to data on Greenhouse Gas Emissions research through the Research Data Australia portal.</td>
</tr>
<tr>
<td>Biodiversity (QUT)</td>
<td>- The key achievements of this project are:</td>
</tr>
<tr>
<td></td>
<td>- The development of software to enable the management and publishing of audio data collections and metadata from acoustic sensors used for measuring environmental health within the ecological research domain.</td>
</tr>
<tr>
<td></td>
<td>- Development of system architecture and suitable workflows for publishing research data information to Research Data Australia.</td>
</tr>
<tr>
<td></td>
<td>- A selection of over 11 records described and contributed to Research Data Australia providing greater exposure to data on Biodiversity research through the Research Data Australia portal.</td>
</tr>
<tr>
<td>B150 Big Jam (QUT)</td>
<td>The key achievements of this project are:</td>
</tr>
<tr>
<td></td>
<td>- The development of software to support the management of multimedia objects and related metadata for musical recordings of the Queensland’s Q150 Big Jam Live Music Festival (2009).</td>
</tr>
<tr>
<td></td>
<td>- Development of system architecture and suitable workflows for publishing research data information to Research Data Australia.</td>
</tr>
<tr>
<td></td>
<td>- A selection of 11 records described and contributed to Research Data Australia providing greater exposure to data through the Research Data Australia portal.</td>
</tr>
<tr>
<td>Smart Water (Griffith University)</td>
<td>The key achievements of this project are:</td>
</tr>
<tr>
<td></td>
<td>- The development of software to support the automated capture of water data.</td>
</tr>
<tr>
<td></td>
<td>- Collection descriptions as well as associated party and activity descriptions harvested via an OAI-PMH interface and made visible through the Australian Research Data Commons (ARDC).</td>
</tr>
</tbody>
</table>

30 September 2012

ands.org.au
<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ands.org.au</td>
<td></td>
</tr>
<tr>
<td>Project</td>
<td>Project Outcomes</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Automated measurement of the responses of wildlife populations to climate change (Flinders University)</td>
<td>The key achievements this project produced are:</td>
</tr>
<tr>
<td></td>
<td>▪ operational system for cleansing sleepy lizard behavioural ecology data collected by Professor Mike Bull and his research team;</td>
</tr>
<tr>
<td></td>
<td>▪ draft research management policies produced and moved through Flinders University’s policy ratification processes;</td>
</tr>
<tr>
<td></td>
<td>▪ operational institutional research data metadata store (ReDBox).</td>
</tr>
<tr>
<td>Melbourne Neuropsychiatry Centre (MNC) Bioinformatics Development Project (University of Melbourne)</td>
<td>The key achievements this project produced are:</td>
</tr>
<tr>
<td></td>
<td>▪ to increase and standardise the information presented on the web in the internal web tool, Business Glossary, and publically in Research Data Australia;</td>
</tr>
<tr>
<td></td>
<td>▪ to publish information about the data in 20 Biogrid databases to Research Data Australia;</td>
</tr>
<tr>
<td></td>
<td>▪ to enhance information in other Biogrid databases to ANDS standards so that publication will be simple when and if the data custodian chooses to publish;</td>
</tr>
<tr>
<td></td>
<td>▪ to export metadata directly to Research Data Australia;</td>
</tr>
<tr>
<td></td>
<td>▪ to increase the likelihood of researchers with similar interests finding out about Biogrid datasets and thereby to increase collaboration.</td>
</tr>
<tr>
<td>Youth Research Centre’s Life Patterns Project: Longitudinal qualitative and quantitative survey data capture and reuse (University of Melbourne)</td>
<td>The key achievements this project produced are:</td>
</tr>
<tr>
<td></td>
<td>▪ Implement and populate the Online Heritage Resource Manager (OHRM) software to support local research data management. The OHRM content will be released as a public access website to inform and encourage collaboration with local researchers.</td>
</tr>
<tr>
<td></td>
<td>▪ Export from the YRC OHRM to the University of Melbourne’s Research Data Registry, and then harvest to Research Data Australia</td>
</tr>
<tr>
<td></td>
<td>▪ Secure ongoing ARC funding, based directly on the success of the ANDS Data Capture project, to build over three years a national database of young Australians at risk of school non-completion.</td>
</tr>
<tr>
<td>Video data in the Social Sciences. Optimising Metadata Capture,</td>
<td>The key achievements this project produced are:</td>
</tr>
<tr>
<td></td>
<td>▪ apps.plantaccelerator.org.au</td>
</tr>
<tr>
<td></td>
<td>▪ the deposit of software developed in the project into open source.</td>
</tr>
<tr>
<td>Project</td>
<td>Project Outcomes</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Data Sharing Procedures and Long-term Reuse (University of Melbourne)   | • Successfully exported metadata from their local system (based on NetXposure) and published it to Research Data Australia, through the local University of Melbourne Research Data Registry  
• Explored metadata issues associated with large-scale audio-visual repositories  
• Established workflows to enable efficient generation of metadata  
• Provided a solid base for other University of Melbourne video based research projects, including metadata tags and descriptions.                                                                                                                                                                                                                                                                                                                                                       |
| Federated Neuroimaging Collections in the National Data Commons (University of Melbourne) | The key achievements this project produced are:  
• Exporting metadata from local Mediaflux-based system (called DaRIS) to Research Data Australia through local Research Data Registry  
• Mapping local DaRIS metadata schema to ANDS RIF-CS metadata schema  
• Improvements to local metadata structure and quality implemented                                                                                                                                                                                                                                                                                                                                                           |
| Humanities and Social Science Research Data at the University of Melbourne (University of Melbourne) | The key achievement this project produced are:  
• Successful demonstration of tools, services and workflows to disseminate information about Humanities and Social Science research data collections to University of Melbourne’s Research Data Registry and Research Data Australia  
• Refactoring the eSRC Online Heritage Resource Manager (OHRM) data management tool  
• Developing tools to ingest data into the OHRM software, and export metadata about multiple data collections held within the OHRM  
• Testing the tools developed on two flagship OHRM projects; The Australian Women’s Register, and the Encyclopaedia of Australian Science.                                                                                                                                                                                                                                                                                       |
| Capture of Complex Data to Support Clinical Research in Cardiovascular and Neurological Medicine (University of Melbourne) | The key achievements this project produced are:  
• the implementation of a new DaRIS repository (extending Mediaflux) for managing research data and the customisation and reuse of functionality developed through other University of Melbourne ANDS Data Capture projects.                                                                                                                                                                                                                                                                                                                                                       |
<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
</table>
| Founders and Survivors Project (University of Melbourne) | The Founders and Survivors Project brings together records relating to 73,000 convicts transported to Tasmania in the 19th century and their descendants. The resulting population database is of national and international significance for historical, demographic and population health researchers. The key achievements this data capture project produced are:  
- To integrate ANDS RIF-CS descriptive metadata within the Founders and Survivors database repository development, improving intra-project communications  
- To describe over 300 collections within the project for export to University of Melbourne Research Data Registry and Research Data Australia  
- To provide a comprehensive set of convict ship index collections, including convict names and basic identity data  
- Development of an XML workflow toolkit to ingest data into the population database, and describe derived datasets for export out to Research Data Australia |
<p>| Research Data Management of the Monash Weather &amp; Climate Program (Monash University) | The project supported Monash University existing weather and climate research and enabled Monash University Weather&amp;Climate (MW&amp;C) researchers to contribute to the work of evaluating the newly deployed Australian ACCESS climate model. The new developed solution also provided the MW&amp;C researchers with the ability to conduct more effective research based on exploiting existing data sets, creation and storage of new data sets and better collaborative research opportunities. |
| Biomedical Data Platform (Monash University) | The project provided protein crystallographers with the ability to conduct more effective research based on exploiting existing data sets, creation and storage of new data sets and better collaborative research opportunities. |
| Tools for curating and publishing | This project has developed the Squirrel solution to integrate with a |</p>
<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>research data in the form of media collections (Monash University)</td>
<td>large institutional research data store (LaRDS at Monash University). Squirrel has provided researchers across a wide range of disciplines with the ability to conduct more effective research based on exploiting existing data sets, creation and storage of new data sets and better collaborative research opportunities.</td>
</tr>
<tr>
<td>Capture and publication of Australian ecosystem data from a network of measurement sites (Monash University)</td>
<td>The solution provided by the ecosystem project has enhanced the research process and provided new research opportunities. The solution has benefited the OzFlux Community of Researchers by providing them with the ability to perform research at the national level using a single data management platform that brings together all Australian institutions doing Ecosystem research. The standardisation of data formatting has led to improved research efficiency, by reducing the time and effort spent on pre-analysis data preparation, thereby allowing the researchers to focus on the analysis of this data to derive research outcomes.</td>
</tr>
<tr>
<td>Capture and publication of data on the history of adoption (Monash University)</td>
<td>The key achievements this project produced are:</td>
</tr>
<tr>
<td></td>
<td>- The process of capturing stories and associated metadata from the website submission form and storing them in a Data Management system.</td>
</tr>
<tr>
<td></td>
<td>- The process of generating a story web page with attached story transcript, metadata files in MODS and DC formats and search tags.</td>
</tr>
<tr>
<td></td>
<td>- The process of ‘publishing’ a story to ARROW, the Monash Library public repository.</td>
</tr>
<tr>
<td>Data Publication to Interferome (Monash University)</td>
<td>The project addressed current shortcoming in collaborations and sharing of immunology and genomics data. The availability of the data and their corresponding collections in Research Data Australia will enable access to all researchers and facilitate hypothesis generation and novel biological discoveries that would not otherwise be possible.</td>
</tr>
<tr>
<td>Data Capture from High Performance Computing Multi-User Environments (RMIT University)</td>
<td>The key achievements this project produced are:</td>
</tr>
<tr>
<td></td>
<td>- RMIT is a very high user of physics and chemistry simulation packages such as VASP, CRYSTAL, SIESTA, and GULP running on a number of High Performance Computing (HPC) facilities. The developed software HPCTardis from this ANDS funded project would interface with the HPC facilities and would assist researchers in collecting, managing and storing their data. The software would also facilitate subsequent retrieval and reuse of generated data that would enable the researchers to curate the data generated from simulation runs. The curation process creates domain specific metadata registry and also provides connector software for publishing metadata to Research Data Australia portal, which facilitates research data be discovered</td>
</tr>
</tbody>
</table>
The software HPCTardis is made available to a public accessible space (GoogleCode), which may attract interests from e-research community. HPCTardis is a modified and extended version of myTardis. HPCTardis has the following features:

- HPCTardis is equipped with newly developed protocol to create experiment automatically in the HPCTardis web portal. The protocol also transfers datasets from the HPC Facilities to the institutional repository (HPCTardis store).
- The extended functionalities in HPCTardis are capable of communicating with Unix servers seamlessly using Unix scripts.
- HPCTardis is equipped with functionalities to extract metadata from four simulation packages such as SIESTA, CRYSTAL, VASP and GULP.
- HPCTardis is developed with various functions, which is capable of producing ANDS specific metadata. The functions can also generate RIF-CS dynamically with related party, activity and collection records that can be harvested automatically using OAI-PMH.

- A selection of 28 records described and contributed to Research Data Australia providing greater exposure to data on Theoretical and Computational Chemistry and Condensed Matter Physics research through the Research Data Australia portal.
- A poster was displayed and a talk was given in eResearch Australasia conference 2011.

<table>
<thead>
<tr>
<th>Project</th>
<th>Project Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMSS RLI Metadata Capture and Publication (La Trobe University)</td>
<td>The project leveraged the existing ANDS services to support Australian researcher’s use of instrumentation at the La Trobe Centre for Materials and Surface Science facility (CMSS). The major outcome from the researchers' point of view is that the system provides a facility to store their datasets, describe them, and easily share them online, with persistent identifiers so that they can be cited conveniently. Furthermore, the new developed system has also provided a platform for the CMSS to publish data acquired from several standard materials. Finally, the system is also able to expose these datasets to the wider world through the Research Data Australia portal.</td>
</tr>
<tr>
<td>Enhanced Metadata Capture for Sustainable Management, Sharing and Reuse of APN Histopathology Research Data (University of Melbourne)</td>
<td>The project aims to link metadata seamlessly with images and other investigational information obtained during histopathological investigations of mice that are models for disease, as part of an Australia-wide initiative to enhance biomedical research. The build phase of the project resulted in a Microsoft Access-based prototype Case Management System (CMS) being deployed at the Histology and Organ Pathology Service (HOPS) in 2010, while a more robust open source web version was implemented in early 2011.</td>
</tr>
</tbody>
</table>
2.2.4 Program Highlights, Issues and Breakthroughs

The main highlight of the program has been the completion of around half of the projects funded under the fast start and EOI processes, with the result that a number of pieces of data management software have been made available, as well as a substantial increase in the amount of data that is under management or described and available through Research Data Australia. The MyTARDIS software has been a particularly popular reuse candidate.

The main issue has been the speed at which projects have been completed. We had expected that most projects would take less than a year to complete, but this has not been the case. There have been a variety of reasons for the delays, with none predominating. They have included staff turnover, failure to manage the project timelines effectively, failure to understand ANDS’ requirements and poor internal communication. In some cases delays in the ANDS processes have also slowed work down, although we believe this to be a rare occurrence.

2.2.5 Program Learnings

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

Learning: Projects need more advice on what they will need to do up front, and more guidance on what they sign up for to ensure they don’t over commit. Additionally, they need to be watched carefully to ensure they are making steady progress. We have applied these principles to the metadata stores project processes to ensure that they finish on time.

2.3 Metadata Stores

2.3.1 Overview of program

Information that can be held about data (often called metadata) can be grouped into four categories. The first is information for discovery, and is primarily held at the level of a collection. This consists of the range of pieces of information that will assist in the discovery of the collection. The second is information for determination of value (also primarily at collection level). This includes information such as the name of the researcher, institution or funding program that might help a potential user to decide whether they want to access the data. The third is information for access that might be a direct link to the data objects (stored elsewhere, such as on national and institutional data stores), both at collection and possibly object level, or contact details for where to source the data. The fourth kind of information is information for reuse, and will include things like reading scales, field names, variables, calibration settings that are needed in order to effectively reuse the data. This will mostly be at object level.
In practice, the distinction between data and metadata can be somewhat arbitrary and depends on the system that is being used to manage the data. If this system is files-oriented, then the metadata will almost always be separately managed in some sort of associated system. If data management system is database-oriented, then much of the metadata will either be attributes of rows and columns for the database tables.

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

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

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

- **Combined Stores**: manage both Collections and Object Metadata for a single institution across a range of disciplines.
- **Collection Stores**: manage the information about data collections within an institution; may also accept feeds from enterprise systems (some of which ANDS has funded), and also feed the ANDS Data Collections Registry.
- **Instrument Stores**: tightly coupled to particular instruments or clusters of instruments. A significant number of these have been developed, not with Metadata Stores funding, but with Data Capture funding. These solutions either feed the ANDS Collections Registry directly (the commonest pattern), or via an institutional Collections store (much less common).
- **For some disciplines**, there are well-established international practices for managing data and metadata, as well as associated software. These Discipline Store solutions might be deployed within institutions or at national or international data centres. ANDS might fund some pipes from instances of these to institutional Collection stores.

As well as these different kinds of metadata stores, the data itself needs to be stored somewhere. This might be a local store (either just associated with an instrument or institutionally supported), one of the offerings

30 September 2012

ands.org.au
that might be made available through RDSI, or an international disciplined-focussed data store (such as the PDB or EMBL/EBI).

Based on our existing engagements with ANDS partners, the most common implementation pattern at an institutional level is a Collection Store to support Seeding the Commons funded projects, combined with one or more Instrument Stores associated with Data Capture funded projects. In this pattern, the Collections descriptions for the Data Capture data are usually fed directly to ANDS rather than via the Collection Store.

### 2.3.2 Outline of projects

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

ANDS has also funded two Collection Stores from the Metadata Stores program. The VITRO solution was developed by Melbourne University (based on the VITRO software and ontology from Cornell) and was deployed by QUT and Griffith University as the basis for their Research Metadata Store Hub. The ReDBox solution is being developed by QCIF and was initially deployed by the University of Newcastle. Both solutions are being used to support Seeding the Commons and Data Capture funded projects at those universities.

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

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

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

1. Larger institutions were invited to participate, rather than all institutions (on the assumption that they had more research data to manage).
2. All institutions who were offered funding were offered the same amount (on the assumption that the costs to deploy were likely to be roughly equivalent for each institution).
3. Institutions that had already received Metadata Stores funding received an amount discounted by the existing funding – in one case this meant no funding went to one of the Data Capture institutions.
4. Funding could be used to enhance existing solutions, deploy new solutions, or provide improved connections to institutional context.

5. Institutions were expected to demonstrate a whole-of-institution commitment to their metadata infrastructure.

Institutions needed to install a metadata store solution that was integrated with other data infrastructure and provided at least a feed to Research Data Australia. All the projects are therefore similar, and vary in detail or specific requirements, rather than overall direction. As such, the individual projects are not documented here. All but two of the institutions (Curtin University of Technology and University of Wollongong) who were offered this funding took it up and the vast bulk of the projects commenced in the first half of 2012.

The following institutions accepted Metadata Stores funding under these conditions:

- ANU
- CSIRO
- Deakin University
- Flinders University
- Griffith University
- James Cook University
- La Trobe University
- Macquarie University
- QUT
- RMIT
- Swinburne University of Technology
- University of Adelaide
- University of Melbourne
- University of Newcastle
- University of New South Wales
- University of Queensland
- University of South Australia
- University of Sydney
- University of Tasmania
- University of Technology, Sydney
- University of Western Australia
2.3.3  Activity/Deliverables for 2011-12

The ANDS funding for QCIF to augment and support the ReDBox solution has now completed. As part of the wider Metadata Stores deployment, a number of institutions have selected this solution. The remainder have either not yet made a selection by the end of this reporting period, have selected VITRO as their technology platform or decided to build their own systems. Most of these projects are still underway at the end of this reporting period, and their deliverables will be the subject of the next annual report.

The ANSTO/Australian Synchrotron project has now successfully completed and is in production use at both organisations.

In addition to the above projects, there has also been a significant amount of metadata store activity that has been funded/cross-subsidised by projects under the Seeding the Commons and Data Capture programs.

2.3.4  Program Highlights, Issues and Breakthroughs

One of the highlights of the program this year was the decision to provide Metadata Stores funding to 22 institutions. This was very positively received by the recipients, with them providing in-kind or co-investment contributions of more than $2 million, compared to ANDS’ investment of $5 million. It is clear that ANDS’ efforts have led to an increase in awareness across all the institutions with which ANDS has been engaging of the importance of metadata about research data outputs, and the need to make an institutional commitment.

2.3.5  Program Learning

Over the previous reporting period, ANDS considered a number of approaches to decide on the optimum way of funding a broad institutionally based Metadata Stores program. The feedback on all of these was that they were unlikely to get sufficient uptake. The approach that was adopted in late 2011 was received well, with the feedback being that the time was now right for this sort of initiative. This has been a benefit of the multi-program approach used by ANDS. The activities of other programs such as Seeding the Commons and Data Capture have helped to develop relationships with institutions and awareness of the needs of software such as that provided by the Metadata Stores program. The idea of another piece of software being offering in isolation was less attractive to our partners. However the greater awareness of the issues involved in data management that the other programs have engendered meant that more institutions were ready to take the extra work in hand.

2.4 Public Sector Data

2.4.1  Overview of program

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

The Public Sector Data program provides the infrastructure necessary to ensure that feeds of data collection descriptions are made available from a range of public sector agencies. Identified agencies include producers of research data, such as the Bureau of Meteorology (BOM), the Australian Bureau of Statistics (ABS), GeoScience Australia (GA), CSIRO and Departments of Primary Industry (DPI). Owners of data gathering activities and collections, such as the museum and library sectors, which might be possible inputs to other research activities, are also in scope. ANDS also needs to maintain and develop stronger relationships with other organisations with significant data holdings or interest in these areas such as the National Archives Australia (NAA) and the Australian Government Information Management Office (AGIMO), for example. Finally, ANDS explores ways to incorporate public data collected by citizens, through exemplar projects.

The key deliverable from this program is to make existing public sector data resources more discoverable to the research community and to work with federal, state and territory government agencies to improve access to data. Activities will vary across agencies according to their existing infrastructure and the types of data being made available. In all cases there will be a strong preference to have data services exposed using relevant international standards.

The Public Sector Data program was originally allocated a $10m budget in the ARDC Draft Project Plan. During the review phases for ANDS mid 2010 this budget was reduced to $6.45m. This was as a result of discussions with key government agencies in the first quarter of 2010 where they identified that their desire was for capability assisted infrastructure development from ANDS in preference to the provision of funding to undertake the infrastructure development themselves.

### 2.4.2 Outline of projects

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

<table>
<thead>
<tr>
<th>Agency or Institution or Project</th>
<th>Project Status and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EIF003 – Auscope</strong></td>
<td>Final Report submitted. Approval pending</td>
</tr>
<tr>
<td><strong>SISS Deployment</strong></td>
<td><strong>AuScope Discovery Portal</strong></td>
</tr>
<tr>
<td><strong>AuScope in RDA</strong></td>
<td></td>
</tr>
<tr>
<td><strong>EIF024 – Australian Legal Information Institute (AustLII)</strong></td>
<td>Complete. 501 collections contributed to RDA. Potential for further definition in National Collections.</td>
</tr>
<tr>
<td></td>
<td><strong>AustLII</strong></td>
</tr>
<tr>
<td></td>
<td><strong>AustLII in RDA</strong></td>
</tr>
<tr>
<td>Agency or Institution or Project</td>
<td>Project Status and Description</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>ELF041 - PowerHouse Museum (CAMD/MA)</td>
<td>At final report stage. 1057 collections contributed to RDA. Software module for the creation of collections descriptions developed for major collection management software and released into open source for uptake by other vendors. Canvassing and initial contribution of university museum collections to MME. <a href="#">Museum Metadata Exchange</a></td>
</tr>
<tr>
<td>National Criminal Justice Research Data Network</td>
<td>Complete. Data hosted on ADA. Discussions underway with ADA to obtain a metadata feed from their archive to RDA. <a href="#">NCJRDN portal</a></td>
</tr>
<tr>
<td>GeoSciences Australia</td>
<td>Initial feed of 7,000 collections complete. Ongoing work through this period to add ANZSC:FOR codes to collections and add further 1700 collections to initial feed.</td>
</tr>
<tr>
<td>Australian Institute of Health and Welfare</td>
<td>Complete. 17 significant collections contributed to RDA.</td>
</tr>
<tr>
<td>Australian Bureau of Statistics</td>
<td>Code developed for ABS to harvest time series records to RDA. ABS looking to secure funding to implement. Investigations to add ABS.stat as a ‘see also’ service in RDA underway. Participation in a demonstrator project to make census data available for AURIN.</td>
</tr>
<tr>
<td>Bureau of Meteorology</td>
<td>Engagement here is pending an enterprise reorganisation of their data and metadata. However there is potential in the next period to expose, via RDA, some of the urban water data they hold.</td>
</tr>
<tr>
<td>Murray Darling Basin Authority</td>
<td>Engagement commenced. Consultation of data holdings and metadata framework underway. Initial exposure of image collection pending.</td>
</tr>
<tr>
<td>Public Records Office of Victoria</td>
<td>Contract executed. Project plan underway. Intention is to expose archives collections from PROV, State Records NSW, State Archives of Queensland and NAA initially. Working with Council of Australasian Archives and Records Authorities to ultimately expose all Australian archives data.</td>
</tr>
</tbody>
</table>
### 2.4.3 Activity/Deliverables for 2011-12

A number of contracted activities were finalised during this period. AustLII completed their contract delivering 501 collections – well over the target agreed. There is potential for further work identifying some national collections of significance to legal research. The Museum Metadata Exchange project is now at final report stage. The portal was highly commended at the annual international “Museums & the Web” conference in the US in April 2012. It came second to the Collections at the Metropolitan Museum of Art, New York, so was in good company. The exchange has delivered 1057 collections to Research Data Australia and requests to universities to include a feed of their museum collections to MME in their exposure of data is now a part of the Better Data programme. Flinders University is in discussion with the Powerhouse on establishing a feed of their metadata.

AuScope has now submitted its final report and, pending implementation of the NLA party infrastructure in SISS, deliverables will be complete. The objective of this project was to make public sector spatial data Services and Collections discoverable and usable by the Australian research community. The objectives were met in a sustainable manner by ensuring that the targeted public sector agencies themselves were involved in the development and subsequent support of the developed infrastructure. The project team intends to continue to publish collection and service descriptions as newly proposed projects and national initiatives are undertaken. The original project planned and delivered on engagements with multiple organisations. The last year has narrowed focus and spent effort primarily on two organisations with data of national importance - the Bureau of Meteorology (BOM) and Geoscience Australia (GA). As a result of this project, and with the collaboration of ANDS staff, both organisations have now defined data management as a mission critical operational goal. GA has over 8700 Collections described in RDA, with most having direct links to data. Satellite imagery continues to be migrated to the National Computational Infrastructure (NCI) where it will soon be publicly accessible. The project will also collaborate with ANDS to develop metadata for ASTER collections and subsequently harvest into RDA. BOM has put in place the infrastructure to expose GeoFabric data and will use this same reference implementation for the National Environment Information Infrastructure (NEII).

Engagements with other agencies include Australian Institute of Health and Welfare (AIHW), Australian Antarctic Division (AAD) and Murray Darling Basin authority (MDBA). An automated feed has been established with AIHW and 17 significant collections have been published to RDA with the promise of more.

<table>
<thead>
<tr>
<th>Agency or Institution or Project</th>
<th>Project Status and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlas of Living Australia</td>
<td>Engagement underway. 118 collections ingested into test area (demo) in preparation for transfer to production. Investigations into harvesting richer metadata from 97 ALA geoserver records to RIF-CS also in progress.</td>
</tr>
<tr>
<td>Australian Antarctic Division</td>
<td>Engagement underway. Test of record feed complete. Ingest of collections about to commence.</td>
</tr>
</tbody>
</table>

Table 3: Public Sector Data projects
to come. AAD is in the process of checking the transformation of records into RIF-CS and once complete the majority of their considerable collection of data will be published, a significant amount not available via AODN. MDBA is testing the process by publishing some of their large image collection. There is more work to be done on the IT infrastructure in MDBA before other data collections can be publically exposed but progress is being made.

The team have engaged with Atlas of Living Australia (ALA) and are working with them and developers from IMOS and CSIRO to devise richer transformation from ISO 19115 to RIF-CS via GeoNetwork which is commonly used in the geo-spatial environment.

PSD had also had a lengthy engagement with Australian Bureau of Statistics (ABS). ANDS has provided the mapping and code for the transformation of time series data from AGLS to RIF-CS for exposure in RDA. This is yet to be implemented in ABS as they are seeking funds to enable it. ANDS is also investigating the feasibility of implementing ABS.stat as a see also service and PSD staff have written code to enable that. Finally ANDS PSD staff are working with ABS and AURIN staff to expose the 2011 census data.

PSD is also working with Bureau of Meteorology (BoM). Originally the data was going to be exposed via the ANDS funded SISS project. However BoM did not have the data ready for the transformation and have embarked on an enterprise approach to the publication of their data. ANDS engagement has helped BoM in developing an enterprise wide approach to its data and associated metadata.

PSD has also funded a contract with the Public Records Office of Victoria (PROV) to establish feeds from archives to Research Data Australia. Working with the Council of Australasian Archives and Records Authorities, the aim is to initially expose metadata from PROV, State Records NSW, State Archives of Queensland and National Archives of Australia with other archives authorities to follow.

### 2.4.4 Program Highlights, Issues and Breakthroughs

PSD has continued its strong relationship with Geosciences Australia and there has been continued improvement to the feeds of metadata. This period saw the application of ANZSRC-FOR codes to the collections. Another highlight has been the establishment of connections between public sector data and research and research data. Feeds from university museum collections into MME will strengthen that links and bring together similar data from a single discovery point. The international acknowledgement of the quality of the Museum Metadata Exchange was also a highlight to this work. Some of the data collections are exposed for the first time and with the software module now available for the museum management systems, this data will be described in collections for the first time. In a similar vein, the establishment of the funded project with PROV has the potential for enabling the discovery of all Australian archives data from a single discovery point for the first time.

Public sector data, while a long established capability, is not always in a state that facilitates easy publication. Up until relatively recently data was largely provided on demand which meant that the descriptions were not always as comprehensive as would be in place for ‘end user’ browsing. Consequently the PSD team have experienced varying levels of completeness and organisation in metadata standards in the agencies with whom we have dealt. Progress has been made in enhancing an awareness of the need for greater access.
Unfortunately requirements for budgetary constraint coincided with the discovery of a requirement for work to be completed on improving description for publication, and this work has sometimes been ‘non-core business’ that was discontinued. However despite this, the increased awareness for the demand for public sector data and the benefit of publications has gained some traction and ANDS work has laid a foundation for future progress.

2.4.5 Program Learnings

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

2.5 ARDC Core Infrastructure

2.5.1 Overview of program

Researchers and research organisations are increasingly expected to make their data public. This allows verification of research claims and the building of innovation upon previous work. Similarly, public funders of research are increasingly requiring public access to the inputs and outputs of research. Research assessment frameworks are also moving towards ways of acknowledging the publication of data as a research output.

The ARDC Core program is building out the national infrastructure services that enable research data to be published and discovered through a program of development of software utilities, both in-house and with partners, which enable organisations and individuals to publicise and share the collections they hold.

2.5.2 Outline of projects

Register My Data

ANDS Register My Data enables the harvesting 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://services.ands.org.au/pages. Register My Data is underpinned by the ANDS Collections Registry.

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

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. It is the face of the ANDS

30 September 2012

ands.org.au
Register My Data service. Research Data Australia includes discovery tools and spatial coverage display. ANDS has implemented a number of significant enhancements to Research Data Australia in 2011-12 (more details in Section 10.3.4) and will continue to enhance it in 2012-13.

Publish My Data

Publish My Data allows individuals to manually register a collection description and to obtain a persistent identifier for their collection. Collection descriptions are stored in the ANDS Collections Registry and are publicly discoverable through Research Data Australia. To be registered, collections must be accessible online.

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

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

Identify My Data

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

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

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

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

- software developers needing to build software which assigns persistent identifiers to objects within their application where there is a requirement to assign, update and resolve identifiers.
- individuals wishing to provide persistence for citing materials held on their web site who may want to assign identifiers to individual pages, papers, or a web site
• institutions acting as an authoritative source of information about people, organisations or community vocabularies where globally unique identifiers need to be assigned to individual records

• institutional repositories not wishing to manage their own identifier service who may wish to assign persistent identifiers to their holdings.

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

DOI Service – Cite My Data

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

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

The ANDS implementation of DOI will complement the existing ANDS Handle-based persistent identifier services.

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

ANDS has developed prototype minting software for Datacite, and contributed to the development of a minimum metadata set. During 2011 has released a pilot service Cite My Data, which allows Australian research organisations and data centres to allocate citation optimised identifiers (DOI). This is one of the enabling services which will allow data collections to be cited and tracked as first-class objects in the scholarly communications cycle. ANDS will develop this infrastructure and tailor it to research organisations’ needs over the next two years.
Vocabulary Service Project

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

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

ANDS has been working to consult the research community on the appropriate development of a Classify My Data service which would 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.

The first phase of the project will establish infrastructure allowing communities to publish and consume existing vocabularies. The functions offered in this phase include:

- Upload vocabularies (GUI and web services)
- Retire vocabularies
- Update vocabularies (GUI and web services)
- Retrieve vocabularies (via web services)
- Download vocabularies
- Browse vocabularies
- Search vocabularies

The functionality to be provided in subsequent phases is still being investigated.

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

Party Infrastructure Project

The objective of this project is to improve the discovery of research data and research publications by linking them through common researchers and research groups. ANDS has partnered with the National Library of
Australia to provide infrastructure to allow Australian researchers and research organisations to more easily use an independent party identifier when publishing their information about data collections.

### Figure 4: Party Infrastructure workflows

Having a persistent identifier assigned to researchers enables different institutional systems 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.

Figure 4 shows the workflows for content providers using the NLA Party Infrastructure.

### Location Infrastructure Project

An important goal of the Australian Research Data Commons is to enable cross-disciplinary discovery of related research data, and spatial location is a vital linkage mechanism in this process. The value of the data commons will be increased if the dataset descriptions include spatial coverage data encoded as geographical points or polygons rather than just text. ANDS' vision for a data commons would see non-GIS-experts from arts, humanities, and science able to enrich their discipline specific data with standardised spatial information.
Achieving this goal requires the establishment of a robust national infrastructure that would allow place names to be validated by both individuals and software systems against an Australian gazetteer service in an efficient manner. There will need to be distributed sources of gazetteer data, depending on jurisdiction, feature types, temporal coverage and language. A comprehensive national service will need to provide interoperable query services across these sources, irrespective of their differing construction.

This infrastructure is intended to increase the amount and quality of spatially-marked-up research data. This will enable new kinds of research and innovation based on new data linkage and data merging opportunities. The infrastructure aims to unlock significant innovation and productivity. It will bring benefits well beyond the research and innovations sector.

ANDS is partnering with Geoscience Australia (GA) to develop and run an online Gazetteer location service. This project seeks to provide 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 Office of Spatial Data Management (OSDM) is encompassed within GA and coordinates the implementation of Australian Government policy on spatial data access.

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. GA will also undertake all software development activities leading to the deployment of human usable web services together with machine to machine services.

The specific deliverables will include:

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

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

**Activity Infrastructure Project**

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

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

In the interim, these records have been ingested as Activity records in to the ANDS Registry, and are available for linking to collection records by ANDS participants, which will increase and add value to the mesh of navigable research records discoverable through Research Data Australia.
2.5.3 Activity/Deliverables for 2011-12

The initial ANDS utility services were launched into production during 2009-10:

- Research Data Australia
- Register My Data
- Identify My Data
- Publish My Data

In 2010-11 ANDS established a commitment to continual improvement balanced with operational reliability. In 2011 the systems began to provide significant support for manual record creation in support of the ANDS Seeding the Commons and Data Capture projects.

The year 2012 saw two considerable upgrades to the ANDS online services:

- Launch of the “Cite My Data” citation identifier infrastructure
- Totally new look and feel and back end to the Research Data Australia site with new information architecture and graphic user interface, and search index

Over the year, new features have also been gradually added to enable data discovery, such as automated record linking, citation exports, full data export in Dublin Core, enhanced harvesting, subject browsing, contributor home pages, RSS and Twitter subscription, global search with DataCite, filtering by license type, and a new simpler URL, <researchdata.ands.org.au>.

Infrastructure and software design approaches have been optimised to allow a much more flexible and dynamic software release cycle. ANDS now uses the NeCTAR National Servers Program to host some of the ANDS Online Services.

The Terrestrial Eco-systems Research Network has adopted the ANDS portal and registry software stack to power its own data federation.

ANDS continues to contribute to the international consortium DataCite, and now provides citation optimised identifiers (Digital Object Identifiers) for use by Australian e-research facilities and data centres. Twelve Australian research organisations, such as CSIRO, are now using the service.

The Location Infrastructure phase two project is near completion. Sustainability of the project is being coordinated by the Committee for Geographical Names of Australasia (CGNA). ANDS has been invited by CGNA to be part of the strategic planning and future directions of the service. ANDS and the Office of Spatial Policy (OSP) combined to upgrade the Gazetteer to address the issues of currency, discoverability and accessibility; and in particular a machine-to-machine capability sought by the research community. The project has will complete in early 2012-13 with the delivery of the following:

- a web-based mapping service with the ability to graphically define search areas and display and download results;
- machine-to-machine functionality;
- an online, near real time update capability for custodians;
compliance with current international best practice guidelines and standards for gazetteers (WFS-G);

- support for the integration of other Australian gazetteers including the Australian Antarctic Gazetteer and the Maritime Gazetteer into a single search environment.


The new Gazetteer infrastructure makes it much easier to get definitive and standardised location information both on the web and integrated programmatically into research applications, tools, and data archives. It makes it easier to “spatially enrich” research data, to connect with data from the same location, and to address complex cross-disciplinary research questions based on location.

In the last twelve months, ANDS has initiated the NLA party infrastructure rollout project and actively promoted the infrastructure for identifying the researchers. This project has led to following outcomes:

- At the time of writing this report, eleven of the top Australian universities have registered as contributors to the NLA party infrastructure, and they have exercised the workflow for integrating NLA identifiers to the process of publishing open research data. This integration enables them to assign unique identifiers to the scholarly authors and enables researchers to receive credits from the reuse of their research data.

- Eighteen Australian universities signed a contract with ANDS to develop metadata management systems that includes the technical capacity to create and maintain unique identifiers for researchers. Another four universities are evaluating similar engagements and the prospect of developing similar projects.

In addition, ANDS has pursued integration with international identity systems including Thomson Reuters’ ResearcherID and the Open Researcher and Contributor ID (ORCID). ANDS is a formal partner in the EU funded ODIN project to explore linkage of identifiers for researchers (ORCID) and datasets (DataCite).

The Vocabulary Support Infrastructure project was de-prioritised for 2011-12. Nevertheless ANDS took the first step of establishing a vocabulary service integrated into the Research Data Australia portal that allows facetted browsing of the ANDS collection according to the Australian and New Zealand Standard Research Codes. ANDS has jointly convened with the Office of Spatial Policy (Commonwealth Dept of Resources, Energy, and Tourism) a Vocabulary Services Technical Working Group with membership from NCRIS facilities and Commonwealth government agencies.

The service desk, change management system, and reporting functionality are now operational as part of a robust national services and infrastructure. Planning has progressed to add a business intelligence reporting capability for more analytical view of the services.

The RIF-CS Advisory Board, constituted and independently chaired by partners and members of the community, continues to give advice to ANDS on changes to the registry interchange format. The Board reviewed and approved a new version of RIF-CS v1.3 in October 2011 and has met regularly during this year to deliberate on a new version for October 2012.

On the following pages are the service usage reports for 2011-12 with historic figures from the previous years:

30 September 2012

ands.org.au
### PUBLIC SYSTEM

<table>
<thead>
<tr>
<th>Service</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handle Service:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handles Minted</td>
<td>4,279</td>
<td>6,323</td>
<td>6,830</td>
</tr>
<tr>
<td>Trusted SW Agreement</td>
<td>10</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>DOI Service (Cite My Data):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOLs Minted (production)</td>
<td>-</td>
<td>-</td>
<td>29</td>
</tr>
<tr>
<td>Registered Publisher Agents</td>
<td>-</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>Registry:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Source Admins</td>
<td>-</td>
<td>46</td>
<td>185</td>
</tr>
<tr>
<td>Provider Org</td>
<td>12</td>
<td>21</td>
<td>59</td>
</tr>
<tr>
<td>Total data source feeds</td>
<td>12</td>
<td>45</td>
<td>122</td>
</tr>
<tr>
<td>Direct</td>
<td>11</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>Automatic</td>
<td>1</td>
<td>15</td>
<td>62</td>
</tr>
<tr>
<td>Publish my Data</td>
<td>51</td>
<td>63</td>
<td>60</td>
</tr>
<tr>
<td>Total Records</td>
<td>2,906</td>
<td>26,746</td>
<td>72,990</td>
</tr>
<tr>
<td>Collections</td>
<td>1,319</td>
<td>23,899</td>
<td>39,957</td>
</tr>
<tr>
<td>Parties</td>
<td>894</td>
<td>1,960</td>
<td>5,473</td>
</tr>
<tr>
<td>Services</td>
<td>3</td>
<td>30</td>
<td>84</td>
</tr>
<tr>
<td>Activities</td>
<td>690</td>
<td>857</td>
<td>27,476</td>
</tr>
</tbody>
</table>

### Research Data Australia:

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Views 2</td>
<td>192,212</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Page Hits</td>
<td>56,659</td>
<td>120,457</td>
<td></td>
</tr>
<tr>
<td>Filtered Page Views</td>
<td>37,527</td>
<td>68,031</td>
<td></td>
</tr>
</tbody>
</table>

### TRIAL SYSTEM

<table>
<thead>
<tr>
<th>Service (Test):</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handles minted</td>
<td>-</td>
<td>15,399</td>
<td>32,937</td>
</tr>
<tr>
<td>Trusted SW Agreement</td>
<td>-</td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>Sandbox:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Source Admins</td>
<td>91</td>
<td>128</td>
<td>137</td>
</tr>
<tr>
<td>Provider Org</td>
<td>60</td>
<td>105</td>
<td>109</td>
</tr>
<tr>
<td>Total Records</td>
<td>33,258</td>
<td>85,881</td>
<td>69,342</td>
</tr>
<tr>
<td>Collections</td>
<td>3,753</td>
<td>32,416</td>
<td>40,233</td>
</tr>
<tr>
<td>Parties</td>
<td>2,392</td>
<td>26,270</td>
<td>7,519</td>
</tr>
<tr>
<td>Services</td>
<td>29</td>
<td>44</td>
<td>114</td>
</tr>
<tr>
<td>Activities</td>
<td>27,084</td>
<td>27,151</td>
<td>21,476</td>
</tr>
</tbody>
</table>

### Table 4: Service usage reports for 2011-12

### NOTES:

1. “Page Views” are all Research Data Australia page views including hits from crawlers and robots (this is the raw data gathered from the web server logs). This method was used in 2009-10.
2. The Google Analytics tool was used to gather page view reports from July 2010 to June 2012.
3. “Page Hits” are Research Data Australia page views cleansed by the Google Analytics tool (thus excluding robots, crawlers etc.).

4. “Filtered Page Views” are Research Data Australia unique page views from Google Analytics tool (filtering out repeat viewings by the same individual in the same session).

5. The DOI Service, Cite My Data, was launched last December 2011. This figure shows DOIs minted since its release.

6. The numbers shown for the total records in Sandbox (trial system) for 2012 are only up to April 2012. The Sandbox environment was decommissioned on 30 April 2012 and records have been migrated to the Production (public system).

![Figure 5: Handles minted](image1)

![Figure 6: User Agreements](image2)
Note: “User Agreements” are agreements for e-research organisations to access ANDS identifier services using the machine to machine interface.

**Figure 7: Provider Organisations**

<table>
<thead>
<tr>
<th>Year</th>
<th>PUBLIC SYSTEM</th>
<th>TRIAL SYSTEM</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>
</tbody>
</table>

**Figure 8: Data Provider Accounts**

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

<table>
<thead>
<tr>
<th>Year</th>
<th>PUBLIC SYSTEM</th>
<th>TRIAL SYSTEM</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>
</tbody>
</table>
Figure 9: ANDS Registry Report

Figure 10: Annual Page Views Report

Note: Page views in 2010 were simply unfiltered logs; from 2011 a filtered page view and unique hit reporting approach was adopted using standardised Google Analytics.
2.5.4 Program Highlights, Issues and Breakthroughs

The launch of the ANDS Cite my Data service has been the major breakthrough of 2012. This launch has been some years in the planning with ANDS playing a foundational part in establishing the international infrastructure to support global identification of datasets (DataCite). Demand for the service has been high despite the long-term nature of the commitment and rewards. Australian research organisations seem to see the potentially significant ongoing benefits of a global system to link researchers with their data output enabling acknowledgement and potentially reward.

The reuse of the ANDS software system by the NCRIS facility TERN has occasioned a change in direction in ANDS software development. The previous design imperative had been to suit ANDS business needs solely. With the advent of other users of the software there is an increased focus on flexibility, extensibility and modularity. During 2011-12 ANDS has laid a foundation for ANDS software to be a specific output of the project rather than simply a tool for providing a service.

The peaking of the Seeding the Commons, Data Capture, Metadata Store, and Public Sector Data projects during 2011-2 has meant that the collection count has doubled and the number of provider organisations has trebled over the year. Usage statistics (above) reflect this uptake and the projections for the coming year are for continued growth. The recent upgrades to ANDS software systems have meant that the system is scaling nicely to cope with growth. The move to NeCTAR application hosting is a step towards server and bandwidth scalability.

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

Again the ANDS partnership approach with long-lived institutions has paid off with the establishment of a national online Gazetteer with GeoScience Australia and the Office of Spatial Policy. The long-term sustainability of the infrastructure is now in the hands of GA and the Committee for Geographical Names of Australasia (CGNA). ANDS has been invited by CGNA to be part of the strategic planning and future directions of the service.

2.5.5 Program Learning

The ANDS project has in general taken an institutional-based engagement approach. This has been reflected in the software systems by a focus on “wholesale” machine to machine connections between ANDS and larger institutional data archives rather than a “retail” approach of individual researchers publishing data with ANDS or minting identifiers. Although our wholesale approach has taken time to establish it is now starting to bear fruit with larger providers like CSIRO providing a feed from their institutional data access infrastructure (which ANDS helped establish). These feeds are established with hundreds or even thousands of data collections descriptions and moreover are dynamically updated from source so that as the organisations data assets grow they are dynamically reflected in Research Data Australia.

The institutional based approach means that we have entered into a partnership with research organisations around Australia to build the infrastructure Australian Research Commons as a cooperative of research data stewards.

30 September 2012

ands.org.au
The dataset identifier services (Cite my Data, etc.) are also taking a similar trajectory with a slow take-off period whilst ANDS works with research organisations to embed the systems, infrastructure, policy, processes, and practices needed for this new research infrastructure component. We envisage that when it does take off as institutional practice, the results will be significant indeed.

## 2.6 ARDC Applications

### 2.6.1 Overview of program

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 intended to leverage the outputs from the other ANDS programs, which have been designed to:

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

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

The goal of the Applications program is to produce compelling demonstrations of the value of having data available for reuse. These demonstrations of value should:

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

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

2.6.3 Activity/Deliverables for 2011-2012
The following Climate Change Adaptation projects have been funded:
- Tropical Data Hub (tropicaldatahub.org);
- Tropical Data Hub - Tools Development;
- Species Distribution Records (spatialecology.jcu.edu.au/Edgar/);
- Climate Change Impacts Downscaling tool;
- Marine Video;
- Health Impacts of Climate Change;
- Climate Change Adaptation Information Hub; and
- Impacts on Ports Infrastructure.

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

With the support from ANDS, the Chromosome 7 working group is the first to develop a viable approach to making the data accessible to scientists. “Our hope would be that all the other chromosome researchers will look at this browser and think, ‘We can use this for our chromosome.’ In fact it could be used as a uniform tool across this initiative internationally,” Prof Ian Smith says. “And there’s no reason why it could not be easily adapted to other fields and organisms.”
The following other projects have been funded:

- Public Open Space modelling;
- SMART Infrastructure dashboard;
- Founders and Survivors;
- CODCD/BMRI demonstrator;
- Marine Virtual Laboratory Information System (blogging at marvlis.blogspot.com.au);
- Ecosystem production in Space and Time (blogging at episat-software.blogspot.com.au);
- TERN/ACEAS Project;
- Urban Climate Modelling; and
- Catchment to Coast.

All but two of the above projects were contracted and commenced during the reporting period. The results of these projects will be described in the next Annual Report.

### 2.6.4 Program Highlights, Issues and Breakthroughs

The main highlight of the Applications program to date has been the successful launch of the Tropical Data Hub. This took place on June 5, and was launched by Minister Evans. Links to coverage of this event can be found at andsapps.blogspot.com.au/2012/06/senator-chris-evans-minister-for.html.

Other highlights have been the appetite within the research sector for this kind of infrastructure activity, and the way in which a number of the projects in the Climate Change Adaptation space have strong connections (both personal and technological) with each other. In the case of this theme, the whole will be bigger than the sum of its parts.

One of the issues confronting the Applications program has been the length of time taken to commence each project. The process from identifying a candidate through a series of discussions to describe the engagement through to contracting has taken up to 12 months in some cases. ANDS has become better at resolving discussions that are not progressing (or where the project is likely to be problematic) and this has helped.

Another challenge has been constructing the balanced portfolio of activities. Early on in the process this was easier as the portfolio had more degrees of freedom. As the process commenced, the criteria that each new project needed to satisfy became more restrictive and demanding. Despite this, we believe the resulting portfolio provides excellent balance.

### 2.6.5 Program Learning

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

2.7 Project Office

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

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

Taking ‘Excellence in operational management’ as its goal, the program worked on streamlining processes and increasing the cohesion of the ANDS group for more effective delivery of service by introducing standard practice, shared tools and cooperative processes. Activities to achieve this have included the consolidation of JIRA and associated workflows for managing projects, engagement and system changes; an overhaul of the ANDS intranet to improve storage and retrieval of internal information with a resultant improvement in information sharing; a makeover for the ANDS website to introduce consistency and reliability to the presentation of ANDS services and information; and a review and improvements to financial reporting, contract management and the production of business plans and annual reporting.

2.7.3 Activity/Deliverables for 2011-12
A major activity has been the establishment of a cohesive approach to communications. A strategy has been developed and common tools and a schedule implemented. Tools have included a planned approach to the use of social media and working with ANDS Capabilities and Research Data teams, uptake in the use of rich media. share has benefitted from the introduction of themes to each issue. A new journalist has been contracted to write the major articles and partners’ contributions have increased allowing a greater number of voices to emerge. ANDS has participated in the launch of the Tropical Data Hub which was successfully executed with the minister, Senator Evans. More such launches are planned for the coming period. Additionally ANDS’ communications has been working with other capabilities on partnered communication efforts. Other events at which ANDS communications have been successfully deployed include eResearch Australasia 2011 and support for ANDS community days and the joint ANDS/NeCTAR developer days.
2.8 Promotion

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

2.8.1 Presentations/attendance at Conferences

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

- Digital Preservation Management Workshop, Melbourne
- eResearch Australasia 2011, Melbourne
- VALA 2012 Conference, Melbourne
- VIVO Community Day, Melbourne
- ReDBox Workshop, Sydney
- ReDBox Community Day, Sydney
- Fundamentals of Data Management for the Humanities, Canberra
- Digital Humanities Australasia 2012, Canberra
- Intersect Showcase 2011, Sydney
- New Horizons for Educational Research, Adelaide
- Information Commissioner Conference, Canberra
- National Forum of eResearch Service Providers, Sydney
- TERN Symposium, Adelaide
- Elsevier Conference on Big Data, E-Science and Science Policy, Canberra
- CSIRO Water Sensor Day, Brisbane
- National Research Infrastructure Capabilities day, Canberra
- Malaria and Open Drug Discovery, Sydney
- ReDS Tinman Workshop, Canberra
- Community Capability Model for Data - Intensive Research Workshop, Melbourne
- Australian Water Recycling Centre of Excellence Workshop, Sydney
- CAIRSS Community Day, Melbourne
- National Linguistics Corpus Launch at Griffith University, Brisbane
- CSIRO Computational and Simulation Sciences and eResearch Annual Conference and Workshops, Melbourne
- Science Communicators Conference, Sydney
- Tropical Data Hub launch at James Cook University, Townsville
- DataCite General Assembly, UK
- Data Curation Centre Conference, UK
- eConcertation Meeting, Lyon, France
- AllHands 2011, York, UK
- Global Research Data Infrastructures (GRDi2020) Workshop, Brussels, Belgium
- 2nd Australia-EU Research Infrastructure Workshop and follow-up Data Workshop, Brussels, Belgium
- iPRES2011 - 8th International Conference on Preservation of Digital Objects, Singapore
- Data Access Interoperability Task Force Meeting, Copenhagen, Denmark
- International Conference on Research Infrastructures (ICRI 2012), Copenhagen, Denmark
- Open Infrastructures for Open Science event, Rome, Italy
- Austrian Research Data Infrastructure meeting, Vienna, Austria
- Knowledge Exchange DAI event, London
- NordBib Conference, Copenhagen, Denmark
- DataCite Summer Meeting, Copenhagen, Denmark
- DataCite Strategy Meeting, Copenhagen, Denmark
- DataCite Summer Meeting and CODATA Data Citation Workshop, San Francisco
- Tools & Data in the Cloud: A series of hands-on Developer Days jointly organized by ANDS and NeCTAR in Melbourne, Brisbane, Perth and Sydney
- Applications Projects Community Days, Sydney and Melbourne
- NSW Metadata Stores and Seeding the Commons roundtables, Sydney
- ANDS Bootcamp, Canberra
- South Australia Community Day, Adelaide

2.8.2 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.

Virtual Events

ANDS has started hosting a variety of free virtual events, to help our partners and communities learn, discuss and exchange ideas, and meet colleagues without even leaving their desk, and they have proven to be incredibly popular.

Virtual events – including webinars, virtual meetings and ‘how-to’ sessions – remove any constraints associated with location, of either the presenter or participants. This has resulted in greater diversity of presenters and topics, which is a great benefit to our local and international audiences and also allows our
communities to hear from and inform international and national perspectives on a wide range of topics and issues.

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

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

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

**Data Citation Roundtables**

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

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

**Tools & Data in the Cloud: A series of hands-on Developer Days jointly organized by ANDS and NeCTAR**

*“A future where researchers and their teams can more easily reuse ANDS data and National eResearch Collaboration Tools and Resources (NeCTAR) tools, as shared national research infrastructure will assure Australia will continue to lead the world in interdisciplinary research...”*

In support of this ANDS and NeCTAR recently held the first of a series of workshops for developers and tech-savvy researchers to demonstrate how data and tools—in the cloud—can be easily reused.

The events not only presented how to use data and tools in the cloud, but also taught participants to launch a tool in to the cloud and then put data into it. Because these tools are in the cloud, they can be demonstrated to researchers at the attendees’ home institutions, thereby demonstrating the time-savings in being able to trial new tools without having to buy a new server or transport a data hard drive across the country to reuse data.

Over 250 people attended the first four events in Melbourne, Brisbane, Perth and Sydney, with similar events in other states to follow in 2012-13.
NSW Metadata Stores roundtables
ANDS—along with the assistance of Vicki Picasso (University of Newcastle) and Peter Sefton (University of Western Sydney), have been running regular metadata stores roundtables so people involved in ANDS Metadata Stores projects can get together and talk collegially.

Applications Projects Community Days
Over 40 participants from 25 institutions come together over two Applications Community Days—in Sydney and Melbourne— to present their projects, discuss a variety of issues, and learn about ANDS and how it can help them achieve research data outcomes.

A number of unexpected connections between projects were revealed such as the potential for some of the Climate Change Adaptation projects to reuse image-viewing technology developed in the medical domain, or usage of the same spatial data sets across multiple projects. Participants also saw how they can draw on each other’s expertise, and many conversations that started during the day have continued beyond the end of the event.

South Australia Community Day
South Australian and Northern Territory Partners came together on 29 June for a Community Day in Adelaide. The event offered an opportunity to share knowledge about research data management, and tips for successful completion of ANDS projects.

The day was a mix of old and the new projects, with a number of projects having completed, many nearing completion, but others including the Metadata Stores projects at the three SA universities and the Charles Darwin University Seeding the Commons project just kicking off.

Sustainability of the research data management effort following completion of funded projects was a recurring theme whilst breakout sessions provided opportunities for informal knowledge sharing.

Informal meetings – Victorian-Tasmanian eResearch/data management community
To further ‘build community’, ANDS commenced an informal monthly meeting—Victorian (Tas.) Informal—that was designed to bring together ANDS partners and guests from IT, Library and Research Offices to discuss eResearch and Data Management. The gathering has a “no minutes, no Powerpoint, just stories” approach, suggested and agreed by the inaugural Informal participants.

This approach frees participants to share both good and bad experiences, and to avoid limits typical when formally representing their institutions. Particularly well attended sessions included chats with UK Digital Curation Centre visitors, and to hear and honour parting reflections from Monash eResearch Centre and

Stefania Riccardi from the Australian Catholic University says:
“At the [Metadata Stores] roundtable I get access to a wealth of “tacit” knowledge that emerges only when people sit together, talk without a script and release more information prompted by objections or the general mood of the audience. Listening and watching Amir [Aryani] talking about parties [NLA Party Infrastructure] made me understand the issues involved much better than any document I have read so far”.

30 September 2012

ands.org.au
VeRSI staff, heading to sunnier climes in Brisbane. Victoria University is collaborating with ANDS on this ongoing-monthly event by hosting the event at their city location.

**ANDS Bootcamp**

The 2012 ANDS Boot Camp was an intensive three-day training program—designed to enhance the eResearch support capacity of those taking a lead role in new ANDS-funded project—which covered: data management and planning, data sharing and access, identifying and describing an institution’s collections, and understanding the legal and regulatory environment for research data.

**Second EU-Australia Workshop on Research Infrastructure**

ANDS, along with a strong contingent of Australian researchers, research infrastructure specialists and research infrastructure funders met with their European Union counterparts at the Second EU-Australia Workshop on Research Infrastructure in Brussels (following the first workshop of this kind in April 2011). There was a special Data Plenary session, jointly chaired by Dr Ross Wilkinson and Dr Peter Wittenburg from the Max Planck Institute for PsychoLinguistics (www.mpi.nl). Data issues and the need to deal with the data explosion were evident in every thematic.

It was clear from the final session that the EU regards their relationship with Australia very highly and wants this workshop series to continue. Indeed, it may be that the next workshop will involve EU participants travelling to Australia in 2013.

**2.8.3 Consultation meetings**

ANDS staff have 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.

**2.8.4 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. Thus three of the four issues in the period have focused on Connecting, Discovering and Reusing our Research Data to complete the spotlight on ANDS’ four transformations, with the last issue of the year themed around Research Data Collections.

30 September 2012
2.8.5 Other activities

ANDS Brochure
In late 2011, the ANDS brochure was updated. It had been over two years since the original brochure was created and it no longer accurately reflected either the function of ANDS or the ANDS branding. The brochure was published in time for eResearch Australasia 2011, and has received positive feedback.

ANDS Website Restructure
The ANDS website (ands.org.au) – our main communication and promotional channel – underwent a restructure in June 2012 to introduce consistency and reliability to the presentation of ANDS services and information. This was done through extensive consultation within ANDS.

Tropical Data Hub Launch
In early June, Minister Chris Evans launched the Tropical Data Hub, at James Cook University. The launch was a great opportunity to directly promote ANDS to the Minister. We also engaged in media relations, which proved to be beneficial in establishing relationships with key media personnel such as the Science Journalist and Education Editor for The Age and a number of Editors of online Science and/or Higher Education publications.

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 first half of 2012. It is worth noting that we have a number of international stakeholders regularly interacting with @andsdata including staff members from JISC (UK), Digital Curation Centre (UK), Heather Piwowar and the University of Newcastle (UK). Tweets that have generated the highest number of click-throughs (meaning engagement with the content) are all promotional tweets about ANDS events, guides and share.

DCC visit
ANDS hosted a visit from a delegation from the Digital Curation Centre (UK) in March 2012.
Sarah Jones and Martin Donnelly from the DCC (dcc.ac.uk) visited ANDS to start a dialogue about sharing ideas, resources and strategies to further data management in both Australia and UK.

Their visit was full of planned sessions for ANDS staff as well as for data managers around Australia, which looked at the DCC and the state-of-play of data management in the UK.
In Melbourne, Sarah and Martin attended the second Victorian eResearch Informal Get-together where participants from Latrobe, Deakin, RMIT, Swinburne, Victoria and Melbourne Universities discussed research data management, intellectual property, archiving and curation. This was a great insight for ANDS partners into the world of data management in the UK.

Sarah and Martin expressed great interest in the strategies ANDS uses to overcome the tyranny of distance to communicate and engage with our community. Also of interest was Research Data Australia going Open Source—it can be downloaded from ands.org.au/resource/techdocs.html

Sarah provides some interesting insights into the visit to ANDS in her blog post: dcc.ac.uk/news/ands-dcc

### 2.9 Risk Management

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

The list of risks is provided in section 10.4. The Risk Register is updated and evaluated once a quarter. In April 2012, ANDS assessed the residual risk level of all 11 risks in the register, taking into account the effect of the risk mitigation strategies that have been put in place.

It was found that 7 of the 11 risks that had been reduced from a rating of high/medium to low in 2011, remained low risks in 2012. These were the risks around political and governance issues as well as relationships with stakeholders and partners. The risk around recruiting and retaining high quality staff also remains low in 2011-12, although it will possibly increase in 2012-13 as ANDS nears the end of its current funding period.

There continues to be medium risk around data providers/federators making their data available and the fact that reusers of research data may not use ANDS services to discover, access and exploit data. However, we have taken measures to mitigate this risk.

Two risks that were lowered from a high to medium rating are unchanged. There is still some confusion about the role of ANDS versus other related service providers in the eResearch sector, but ANDS has been developing and implementing communications strategies to mitigate this risk, including actively engaging with NeCTAR, RDSI and other eResearch organisations to ensure clarity of the different roles. The second risk is that the standards and technologies that ANDS adopts are not adopted more widely. This has been mitigated by seeking international engagements and partnerships to take up standards and technologies favoured by ANDS and share development load. ANDS has also been conducting boot camps, community events and workshops to educate a larger audience on how these technologies are developing in order to increase their readiness to adopt them. Another strategy is to encourage the use of ANDS-developed technologies by other data aggregators, such as Terrestrial Ecosystem Research Network (TERN).
### 3 Progress against milestones

#### 3.1 Data Capture Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>11Q3</strong></td>
<td>All projects underway and project plans delivered to ANDS</td>
<td>50 projects underway, 24 project plans accepted</td>
</tr>
<tr>
<td></td>
<td>Completion of 8 projects</td>
<td>5 projects completed</td>
</tr>
<tr>
<td></td>
<td>Other projects being monitored and assessed as required.</td>
<td>Project monitoring in progress</td>
</tr>
<tr>
<td></td>
<td>Delivery of 70 records to Research Data Australia</td>
<td>14000 records added</td>
</tr>
<tr>
<td></td>
<td>New data management tools available via Open Source</td>
<td>2 data management tools available</td>
</tr>
<tr>
<td></td>
<td>Agreements reached on “Themes”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feeds from previous projects are checked for activity</td>
<td></td>
</tr>
</tbody>
</table>

| **11Q4**       | Completion of 12 projects                                                 | 5 projects completed                                                     |
|                | Other projects being monitored and assessed as required.                 | Project monitoring in progress                                            |
|                | Delivery of 800 records to Research Data Australia                       | 160 records added                                                        |
|                | 15 new data management tools available via Open Source                   | 5 data management tools available                                         |
|                | First version of catalogue of tools available                            | Not done – done in Q1, 2012                                             |

| **12Q1**       | Completion of 10 projects                                                 | 5 projects completed                                                     |
|                | Other projects being monitored and assessed as required.                 | Project monitoring in progress                                            |
|                | Delivery of 300 records to Research Data Australia                       | 83 records added                                                         |
|                | new data management tools available via Open Source                      | 5 data management tools available                                         |
|                | Agreements reached on “Themes”                                            | Not done – no excess funding                                             |
|                | Feeds from previous projects are checked for activity                    | Not done                                                                 |

30 September 2012
<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>12Q2</td>
<td>Completion of 12 projects</td>
<td>14 projects completed</td>
</tr>
<tr>
<td></td>
<td>Other projects being monitored and assessed as required.</td>
<td>Project monitoring in progress</td>
</tr>
<tr>
<td></td>
<td>Delivery of 200 records to Research Data Australia</td>
<td>650 records added</td>
</tr>
<tr>
<td></td>
<td>15 new data management tools available via Open Source</td>
<td>14 data management tools available</td>
</tr>
</tbody>
</table>

### 3.2 Metadata Store Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>11Q3</td>
<td>Work has commenced on adapting/developing an Object Store solution</td>
<td>Object store solution was not pursued.</td>
</tr>
<tr>
<td></td>
<td>Agreements reached with 6 institutions to redeploy an ANDS-funded Collection Store solution</td>
<td>Not achieved in this period, planning was undertaken for a wider roll out of collection stores.</td>
</tr>
<tr>
<td>11Q4</td>
<td>Testing commences for initial version of an Object Store solution</td>
<td>Object store solution was not pursued.</td>
</tr>
<tr>
<td></td>
<td>6 institutions have deployed an ANDS-funded Collection Store solution</td>
<td>Not achieved in this period, planning was undertaken for a wider roll out of collection stores and offer letters were sent to institutions.</td>
</tr>
<tr>
<td>12Q1</td>
<td>An initial deployment of Object Store solution has taken place</td>
<td>Object store solution was not pursued.</td>
</tr>
<tr>
<td></td>
<td>Development/adaptation has commenced for a range of pipes</td>
<td>Pipes were developed as part of individual projects at institutions.</td>
</tr>
<tr>
<td>12Q2</td>
<td>Further deployments of Object Store solution complete</td>
<td>Object store solution was not pursued.</td>
</tr>
<tr>
<td></td>
<td>Initial deployment of pipes (including testing) complete</td>
<td>Pipes were developed as part of individual projects at institutions.</td>
</tr>
</tbody>
</table>
### 3.3 Public Sector Data Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
</table>
| 11Q3           | Environmental scan to determine demand for prioritisation of further engagements  
                 Scheduled engagements with 4 agencies to commence  
                 Engagement with AODN to deliver 5000 data collections | Complete. Primary driver for schedule remains willingness and capacity to engage.  
                 Agencies: ABS, Atlas of Living Australia, AIHW, Murray Darling Basin Authority  
                 AODN data delivered. Continuing feeds now delivering 11,020 data collections of varying size.  
                 Prioritised engagements commenced |
| 11Q4           | Contracted projects to complete delivering 3000 data collections, 8 software items released into open source, 10 deployments of SISS, 1 metadata store and exposure of data from 120 agencies (federal, state and local). | Complete. In addition to their original agreements, GA delivered further 1700 collections, AODN: 6000, AustLII a further 100 and MME: 300 in excess of original undertaking and feed. The SISS deployments have delivered 88 collections with a further 1500+ under review. Associated with the collections the SISS deployment has also delivered 27 records for services associated with the data.  
                 Software released to open source includes the entire suite of SISS software and the MME collections management software. Deployments of SISS are: 4 instances at GA, BoM, OSDM, AuScope Grid, CSIRO/WAGC, VeRSI, Uni of Ballarat, DPI (Vic), DSE (Vic) WAGCE & Co2CRC.  
                 Collections have come from the above deployments plus 18 museums across Australia |
### 3.4 ARDC Core Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>2010-11 Milestones</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>11Q3</td>
<td>Citation Identifier Infrastructure phase one implemented</td>
<td>ANDS DOI service “Cite My Data” launched</td>
</tr>
<tr>
<td></td>
<td>Researcher Identification Infrastructure</td>
<td>National Library <a href="https://ands.org.au">Party Infrastructure</a></td>
</tr>
<tr>
<td>Milestone Date</td>
<td>2010-11 Milestones</td>
<td>Progress</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>11Q4</td>
<td>Registration, Discovery and Data Collection Page Creation Infrastructure phase two complete</td>
<td>Belvedere project completed with total redesign of Research Data Australia Gazetteer Project extended to phase two and completed in Q3 2012</td>
</tr>
<tr>
<td></td>
<td>Place Names Infrastructure complete</td>
<td></td>
</tr>
<tr>
<td>12Q1</td>
<td>Vocabulary Infrastructure complete</td>
<td>Scaled down vocabulary service launched internally to Research Data Australia to allow browsing by ABS classifications</td>
</tr>
<tr>
<td></td>
<td>ABS classifications complete</td>
<td>Updates to Cite my Data completed in May 2012 using feedback from users (xml response formats; URI structure; admin screens)</td>
</tr>
<tr>
<td></td>
<td>Citation Identifier Infrastructure phase two complete</td>
<td></td>
</tr>
<tr>
<td>12Q2</td>
<td>Research Activity Infrastructure complete (ARC and NHMRC)</td>
<td>Slippage; funding body timetables incompatible; project paused</td>
</tr>
<tr>
<td></td>
<td>Data Collection Discovery and Page Publication phase three complete</td>
<td>Major update to Research Data Australia released in July 2012 with extensive new features</td>
</tr>
</tbody>
</table>

### 3.5 ARDC Application Infrastructure

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>10Q3</td>
<td>First demonstrator of value problem selected, and relevant researchers selected</td>
<td>Delayed as a result of delayed choice – see 2.6.5 for discussion. Delayed.</td>
</tr>
<tr>
<td></td>
<td>First candidate institutions for research data champions selected</td>
<td></td>
</tr>
<tr>
<td>10Q4</td>
<td>Development of proposals for first round of research data champions completed</td>
<td>Changed: Initial discussions held with researchers to identify possible opportunities. Delayed as projects were still forming.</td>
</tr>
<tr>
<td></td>
<td>Work for first demonstrator of value scoped</td>
<td>Changed: Hackfests were determined to</td>
</tr>
<tr>
<td>Milestone Date</td>
<td>Milestone</td>
<td>Progress</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>Hackfest date decided</td>
<td>deliver less value so they were no longer prioritised.</td>
</tr>
</tbody>
</table>
| 11Q1           | Work for first demonstrator of value commenced  
Research data champions institution selected and funded, work commenced  
Hackfest organised and held | Delayed as projects were still forming.  
Changed: No longer prioritised. |
| 11Q2           | Hackfest organised and held | Changed: No longer prioritised. |
4 Deviations from the Project Plan

There were a number of deviations to the project plan; the three of most significance are:

- Developing an institutionally focused National Collections program to complement the work of RDSI;
- Developing an International Infrastructure program to respond to the opportunities that emerged out of the Research Data Alliance and associated initiatives; and
- Lower expenditure that budgeted.

National Collections

Preparation for the operational implementation of the National Collections program has taken place in the second half of this period. This program had to exploit the existing relationships with institutions, and an understanding of collections that are important to them, to develop a synergistic relationship with RDSI, and develop a new relationship with RDSI Nodes as they were announced. A team put together a framework for operating which includes a planned approach to how the program will work with partners in cooperation with the other ANDS programs, and activities that ANDS staff will undertake to assist the establishment, description and publication of national collections. The team has been working with the nodes and the RDSI ReDS team to identify candidate collections and plot the work associated with achieving cooperative storage of data that is managed, connected, discoverable and ultimately well reused.

International Infrastructure

ANDS has been asked by DIISRTE to act as the Australian Non-Government Sector (NGS) representative on the emerging Research Data Alliance. This is being brought into existence in partnership with the US and the EU to improve data interoperability and sharing. ANDS has received separate funding for this activity in FY 2012-13, but was resourcing this activity internally in 2011-12. This activity will build on other international engagement – improving data connectivity through DataCite and a new EU funded project called ODIN, engagement with other internationally focused initiatives such as GRDI and EUDAT, as well as many bilateral engagements.

Expenditure Reduction

Actual expenditure was $6.1 million less than budgeted; the main reasons are outlined below.

- This was partly due to lower expenditure on staff than budgeted by approximately $0.6 million:
  - Seeding the Commons: 1 EFT on unpaid sick leave, 1 EFT to be left vacant
  - Data Capture: Delay in start date for 1 EFT
  - National Collections: Delay in recruitment for 2 EFT
- The bulk of the variance (approx. $4.6 million) is a result of delays in contracting and contract invoicing:
Applications ($2.0M): as most contracts are at commencement stage and a couple of contracts were delayed in contracting.

Data Capture ($1.2M): due to a combination of delay in contract payments and some contracts at commencement stage.

Metadata Stores ($0.8M): as most contracts are at commencement stage.

Seeding the Commons ($0.3M): mainly due to a delay in the second round of projects.

Core ($0.3M): mainly due to delay in contracting.

Public Sector Data ($0.5M): funds which were initially planned for 5 eResearch institutions have been reallocated to unallocated funds.

Figure 11: Intended expenditure pattern as at November 2011
Figure 12: Intended expenditure pattern at September 2012
5 Commitments

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

The total value of contracts as at 30th June 2011 was approximately $20.3 million, whereas by 30th June 2012 we had contracted an additional $12.8 million. An amount of approximately $2.1 million worth of projects is institutionally based with Monash University (as Monash is the lead agent, a separate contract could not be signed with it for these projects). As at 30th of June 2012, we had $3.7 million dollars’ worth of commitments, i.e., projects for which contracts were in the process of being negotiation.

<table>
<thead>
<tr>
<th>Contracts Schedule</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Contracts as at 30 June 2011</td>
<td>20,361,297.91</td>
</tr>
<tr>
<td>Total Contracts executed from 1 July 2011 to 30 June 2012</td>
<td>12,833,045.09</td>
</tr>
<tr>
<td>Monash University’s ANDS-funded projects paid via internal transfer</td>
<td>2,100,175.43</td>
</tr>
<tr>
<td>Total Contracts in progress as at 30 June 2012</td>
<td>3,708,586.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>39,003,104.43</strong></td>
</tr>
</tbody>
</table>

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

The following table indicates ANDS expenditure by program for July 2011 to June 2012. Income is shown in the audit statement in Section 9.

**ANDS Expenditure for the Financial Year ended 30th June 2012**

<table>
<thead>
<tr>
<th>Item</th>
<th>2012</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applications</td>
<td>486,020</td>
<td></td>
</tr>
<tr>
<td>Data Capture</td>
<td>343,586</td>
<td></td>
</tr>
<tr>
<td>Monash University Data Capture Project</td>
<td>523,954</td>
<td></td>
</tr>
<tr>
<td>Metadata Stores</td>
<td>211,322</td>
<td></td>
</tr>
<tr>
<td>Monash University Metadata Stores Project</td>
<td>114,000</td>
<td></td>
</tr>
<tr>
<td>Project Office</td>
<td>141,884</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total Salaries</strong></td>
<td><strong>1,820,765</strong></td>
<td></td>
</tr>
<tr>
<td>Operating Expenses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Capture External Infrastructure Contracts</td>
<td>4,681,791</td>
<td></td>
</tr>
<tr>
<td>Data Capture External NeAT Projects</td>
<td>144,785</td>
<td></td>
</tr>
<tr>
<td>Data Capture Internal Program Expenses</td>
<td>11,366</td>
<td>2</td>
</tr>
<tr>
<td>Metadata Stores External Infrastructure Contracts</td>
<td>1,581,968</td>
<td></td>
</tr>
<tr>
<td>Metadata Stores External NeAT Projects</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Metadata Stores Internal Program Expenses</td>
<td>25,293</td>
<td>2</td>
</tr>
<tr>
<td>Public Sector Data External Infrastructure Contracts</td>
<td>701,382</td>
<td></td>
</tr>
<tr>
<td>Public Sector Data External NeAT Projects</td>
<td>56,174</td>
<td></td>
</tr>
<tr>
<td>Public Sector Data Internal Program Expenses</td>
<td>10,947</td>
<td>2</td>
</tr>
<tr>
<td>Applications External Infrastructure Contracts</td>
<td>1,182,615</td>
<td></td>
</tr>
<tr>
<td>Applications Data External NeAT Projects</td>
<td>108,447</td>
<td>2</td>
</tr>
<tr>
<td>Applications Data Internal Program Expenses</td>
<td>60,220</td>
<td></td>
</tr>
<tr>
<td>ARDC Core External Infrastructure Contracts</td>
<td>85,333</td>
<td>2</td>
</tr>
<tr>
<td>ARDC Core Internal Program Expenses</td>
<td>34,460</td>
<td>3</td>
</tr>
<tr>
<td>ARDC Core ANU Program Funding Payment</td>
<td>1,549,955</td>
<td>4</td>
</tr>
<tr>
<td>Project Office Internal Expenses</td>
<td>39,776</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total Operating Expenses</strong></td>
<td><strong>10,274,513</strong></td>
<td></td>
</tr>
<tr>
<td><strong>EIF Total Expenditure</strong></td>
<td><strong>12,095,279</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

1. Half of Executive Director’s and Contract Administration salaries to manage ELF program
2. Institutional engagement costs including travel.
3. Includes DataCite International membership fee 2010-2012.
4. This payment is based on ANU-declared internal costs to run this program.
5. Includes travel expenses with added adjustment of reclass of software licence fees from EIF funds to NCRIS funds.
Figure 14: ANDS Organisational Chart
7 Co-Investment

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

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

ANDS investments at institutions in Seeding the Commons and Data Capture activities have triggered substantial co-investment and post-investment with over $2M of institutional investment made to date, and over $3M of post-project investment. This indicates the extent to which institutions are embedding a research data infrastructure into standard operations.

It is pleasing to note that this additional effort is being continued beyond the life of the projects, showing the importance being placed on research data, and the extent to which institutions are establishing coherent research data infrastructure as part of “business as usual” operations.

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


8 Performance Indicators

8.1 KPI Report

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

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

   **Result:** 59 institutions fed collection descriptions to Research Data Australia along with 45 individually published collections. From these institutions, 122 data source feeds have been setup (62 automatic and 60 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 AustLII project we have exposed public data from Attorney General’s Department and various courts around Australia including the High Court. Through the project with AuScope, engagement has been with Bureau of Meteorology and various state Departments of Primary Industry and Sustainability & Environment
   - National facilities: Australian Animal Health Laboratory (AAHL), Australia Telescope National Facility (ATNF), Australian Synchrotron and research vessels: Southern Surveyor and Aurora Australis
Imaging Facility (NIF), Australian Microscopy & Microanalysis Research Facility (AMMRF), EMBL Australia, Australian National Fabrication Facility, AusBiotech, Research Infrastructure Support Services, Australian Biosecurity Intelligence Network, Australian Social Sciences Data Archive, Australian Access Federation, National Research Network, Pawsey Supercomputing Centre, National eResearch Collaboration Tools and Resources (NeCTAR), Research Data Storage Infrastructure (RDSI).

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

**Result:** 10 – CSIRO, Monash University, University of Melbourne, Queensland University of Technology, Griffith University, University of Wollongong, University of Newcastle, Edith Cowan University, La Trobe University 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:** 47,385 searches since September 21, 2011.

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

**Result:** The target of 300,000 was based on a previous measure of page views that we believe is no longer appropriate.

120,457 page views and 68,031 unique page views (from Google Analytics tool). Page views doubled from last year.

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

**Result:** We have not conducted a similar survey in 2011-12 but instead have taken a number of more specific measures to understand partner sentiment. They include:

- ANDS seeks and captures feedback routinely from every project upon completion.
- ANDS has established a Service Desk to capture feedback from users of ANDS services as well as partners, and a well-documented process is followed to assess and respond to these queries and comments in a timely manner.
- After each event ANDS conducts with the community, an analysis of the effectiveness of the event is sought from the attendees.
- ANDS seeks regular advice from senior research staff within the sector on the effectiveness of the ANDS programs.

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 59 organisations to publish data collection descriptions in either the public or draft systems, up from 49 last year. 51 organisations now publish information to the ANDS production system, up from 21 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: 39957 collections as at June 30, 2012. 34% of these collections were from Queensland Facility for Advanced Bioinformatics (QFAB), 28% came from Australian Ocean Data Network (AODN) and 22% came from Geoscience Australia (GA). Collections from these 3 providers comprise 84% of the total collections in Research Data Australia.

The number of ANDS research data sets increased 67% from June last year.

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

Result: 6859 collections (6830 PIDS handles, 29 DOI handles), 8% higher than June last year.

This is lower than anticipated but the bulk of the work in this space has been around establishing sustainable PID-minting services, especially DOI services, integrated into institutional workflows. There are now 12 registered publisher agents.

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 will again set benchmarks, but also help inform future surveys.

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

Having been in progress for 30 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 2011-12, 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>2009-10</th>
<th>2010-11</th>
<th>2011-12</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Institutions Participation in RDA</td>
<td>NA</td>
<td>21</td>
<td>35</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>43</td>
</tr>
<tr>
<td>Institutional context capture tools</td>
<td>0</td>
<td>6</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Institutional Research Metadata Store</td>
<td>0</td>
<td>9</td>
<td>25</td>
<td>43</td>
</tr>
<tr>
<td>Research Data Provider Participation</td>
<td>NA</td>
<td>9</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Research Data Infrastructure Partners</td>
<td>4</td>
<td></td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Research Data Collections</td>
<td>1,173</td>
<td>26,746</td>
<td>40,811</td>
<td></td>
</tr>
<tr>
<td>Research Data Exploitation Tools</td>
<td></td>
<td>9</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Research Fields of Research Coverage</td>
<td>5</td>
<td>21</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
8.3 Overall Progress

In Section 1.2, ANDS reported 2 major achievements:

1. Establishing the ARDC, and

2. Providing a meeting place that enables research data management for the whole of Australia to be progressed.

As well as substantial progress in meeting another key objective:

3. Populating the ARDC.

There is a strong relationship between our KPIs, established at the start of ANDS, and achieving the objectives of ANDS. We have shown in the previous section that there are other measures that can also be added to help understand progress.

**Establishing Institutional Research Data Infrastructure:** Institutions are closest to researchers that produce and consume research data, so there needs to be a coherent set of institutional research data infrastructure services that interact with national and international services – then a researcher can capture research data, the associated metadata determined by a research data management plan, to store both the data and metadata, to connect the data with its context, to publish that data, and to discover and use research data from others. The measures showing the data collections with a persistent identifier available through Research Data Australia indicate that researchers can publish their data, but makes no distinction between research data outputs and data inputs to researchers. The measures showing research data management planning and associated institutional research data infrastructure such as metadata stores and automated systems for the capture of metadata and connections tools shown in the above table provides this information provide a further evidence of institutional research data infrastructure.

**Increase profile for research data and delivering Australia’s research data advantage:** KPI’s do not directly address this measure – the outputs within Government demonstrate the strength of its commitment to a research data infrastructure, and whilst ANDS has had engagement with Government, it is difficult to demonstrate that role. Internationally, the role the ANDS is playing through the Research Data Alliance as a founding member, as well as many bilateral engagements are best shown through examining the Communications section of this report.

**Maturing the Australian Research Data Commons:** ANDS has refined its objectives in this regard: ANDS wishes to populate the commons with collections that are managed, connected, discoverable, and re-useable, and have in place all infrastructure to support this. Current KPIs particularly focus on registered and discoverable collections, but provide less insight into just how well collections are managed, connected, and re-usable. By way of example, the work done by CSIRO as part of our collaboration with AuScope has led to collections of professionally managed data from the Geological Surveys that are described, made available, and easily integrated with other similar data using a rich set of web services. All research areas are covered, and there are many more institutions providing collection descriptions. There are more connections, and richer collection descriptions. Thus our KPI’s provide measures that are good indicators of maturity, but do not tell the whole story.
9 Audit Statement

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

Australian National Data Service Project - EIF Funding
Statement of Income and Expenditure for the Financial Year Ended 30 June 2012

<table>
<thead>
<tr>
<th>Item</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td></td>
</tr>
<tr>
<td>Grant</td>
<td>-</td>
</tr>
<tr>
<td>Interest</td>
<td>1,455,358</td>
</tr>
<tr>
<td>Other Income</td>
<td>-</td>
</tr>
<tr>
<td>Total Income ((a))</td>
<td>1,455,358</td>
</tr>
<tr>
<td>Expenses</td>
<td></td>
</tr>
<tr>
<td>Salaries</td>
<td>1,820,765</td>
</tr>
<tr>
<td>Non-Salary</td>
<td>10,274,513</td>
</tr>
<tr>
<td>Total Expenses ((b))</td>
<td>12,095,278</td>
</tr>
</tbody>
</table>

Surplus/(Deficit) for the reporting period \((a - b)\) \((10,639,920)\)
Brought forward surplus/(deficit) from 30 June 2011 \((c)\) \((32,482,121)\)
Balance Carried Forward to next Reporting Period \((a - b) + (c)\) \((21,842,201)\)

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

(i) The detailed statement of income and expenditure for the ANDS Establishment Project (shown above) represents a true and fair view of the financial performance for the year ended 30 June 2012.

(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 EDWINA CORNISH
Senior Deputy Vice-Chancellor & DVC, Research
Office of the DVC, Monash University

MR JOEL CHIBERT
Director of Research & Revenue Accounting Services
Research & Revenue Accounting Services, Monash University
10 Appendices

10.1 Confidential Information

There is no confidential information.

10.2 Project Description Detail

10.2.1 Data Capture Fast Start Project Summaries

ANSTO/Australian Synchrotron

These were originally funded as two separate projects. It became clear to ANDS during the early planning stages that there were significant synergies between the two national facilities, as well as an overlap between the researchers they were servicing. Accordingly, ANDS brokered a meeting between the relevant principals of both institutions at the eResearch Australia 2009 Conference to discuss the possibility of a converged project. This meeting was successful, and the project was constituted with a single project committee and project manager.

The Australian Synchrotron operates 9 beamlines producing up to 2TB of experimental data per day across a wide variety of disciplines from protein crystallography, medical (cancer research) through to the conservation and restoration of cultural objects and works of art. Last year (2009) over 500 groups conducted research at the Australian Synchrotron.

The Bragg Institute is the strongest neutron and X-ray scattering group in Australia. The Institute is named as a tribute to the father-and-son team of William and Lawrence Bragg who were jointly awarded the Nobel Prize for Physics in 1915 for pioneering the analysis of crystal structures by means of X-rays. The facility operates a number of neutron beam instruments, including diffractometers, reflectometers and spectrometers.

The objective of this project is to provide services to researchers to manage their experimental data and to provide data search and access to the broader research community. These services will provide better use and reuse of the data. The ultimate aim is to combine these services into a collaboration environment to allow project teams to interact with the instruments and the data. This project will also improve data and metadata capture for the three most mature beamlines at the Australian Synchrotron, the two High-throughput Protein Macro Crystallography (PX1 and PX2) beamlines and the Infrared (IR) beamline. It will develop appropriate metadata schemata, combine metadata from various sources into a richer set of metadata, and set up a metadata store for the metadata from these beamlines that can be harvested by ANDS for publication of the metadata in the ARDC. This metadata store will be designed so that it can later be utilized for other beamlines at the Synchrotron. Furthermore, the project will extend existing facilities for raw data to be transferred into more permanent storage, including preparation for services to transfer data.
from raw storage at the Synchrotron to the ARCS data fabric once this becomes available. For the IR beamline, the project will also develop a reference spectral database of different materials (e.g. paints, proteins, inks) that enables reuse of these data by users of the IR beamline. This database will be hosted at the Synchrotron on a permanent basis.

CSIRO

The scope of the ANDS-CSIRO-ATNF Pulsar Data Management Project is to enable the discovery of, assessment of and access to Pulsar data observed at the Parkes Telescope. The pulsar data observed at the Parkes telescope is publicly available after an embargo period. This project proposes to develop tools required to automate the capture of these data collections, and develop metadata for harvesting into the Australian Research Data commons (ARDC) to increase their discoverability and potential for reuse. It will include:

- Development of data translation toolsets and implementation of metadata standards, to ensure that valuable science data assets are more easily discoverable and more readily reused.
- Establishing infrastructure to allow direct access to datasets allowing researchers to either download subsets of data direct to their desktop or copy complete datasets utilising ARCS data fabric infrastructure to appropriate advanced scientific compute facilities.
- Establishing middleware systems that track and inform researchers about embargoes for the datasets ATNF holds. Where embargoes prevent direct access to data the system will inform the enquirer who owns the data and their contact details.
- The population of the CSIRO repository with metadata describing the pulsar data sets and the subsequent harvest and population of metadata into the Australian Research Data Commons (ARDC) and the Virtual Observatory.

Monash University

Comprehensive Data Management for Microscopy Research Datasets - This solution allows the Faculty of Medicine, Nursing and Health Sciences to generate and capture experimental information and raw data from optical microscopy instruments in a centralised location, rather than the current practise of capturing this on researchers’ labnotes and personal disks. This approach will establish a concept for the mandatory annotation of digital imaging data to maximise the use, reuse and distribution of experimental information within the scientific community.

University of Sydney

This project will construct software to allow wide use and reuse of microscopy images in breast cancer research, generated at the Westmead Institute for Cancer Research (WICR). In principle the audience could include the entire international breast cancer research community. Australia has significant strength in cancer research, and the Australian research population in breast cancer numbers in the thousands, with many more internationally. There is also potential for other cancer research communities to partner in this project, making it a highly significant international resource. There will be several software outcomes from the project:
- Tool to segment and convert segments of extremely large images, make them navigable and downloadable in a size and format suitable for analysis, distribution and reuse
- Web portal for search and discovery of images and sections of images from a repository to
- Tool for harvesting of appropriate image metadata to ANDS collection registry

Taken together, these components will provide a path for images from the scanner, to the immediate Westmead group for analysis, and beyond them in de-identified form for use by the international research community.

The objectives of the project—format conversion, distribution portal, discovery, and identity security—generalise beyond breast cancer to a wide spread of other medical research and large image management. Techniques developed and experience gained in this project will be of broad value for future eResearch development, and particularly useful in the medical area.

**University of NSW**

A new generation of DNA sequencers has recently been installed at UNSW, Southern Cross University and the Australian National University. These instruments can generate DNA sequence data 1000x faster than old technology, and can sequence the genomes of small organisms in a week. This project will establish databases for this DNA sequence information, so that users of the DNA sequencers can access their information in an efficient way. This will centralise DNA sequence data from these facilities, enable collaborative projects and facilitate data sharing. On publishing, research data and associated metadata will be made available for public use, contributing to the Australian Research Data Commons (ARDC).

**University of Melbourne**

This project will be a key part of establishing a national database of mouse pathology to enhance the utilisation of mouse models of disease by Australian researchers. It will investigate enhanced metadata capture facilities for the Histopathology and Organ Pathology Service based at the Department of Anatomy and Cell Biology, The University of Melbourne as part of facilitating the sharing and reuse of mouse pathology research data both now and into the future. The project will address current metadata scalability and sustainability issues associated with the service in order for the Melbourne Histopathology Service to participate in and contribute to emerging research data networks like PODD and ANDS.

**University of Queensland**

This project will develop the data capture and sharing services for coral-reef related data being generated by researchers at the University of Qld Centre for Marine Studies, together with their collaborators in community/volunteer groups (CoralWatch, ReefCheck) and government organizations (EPA, DERM). Together these researchers are monitoring and studying the impact of climate change and human activities on coral reef ecosystems. This project will focus on data associated with reefs located between the Southern Great Barrier Reef and the QLD/NSW border. The data will include the following types of observational data: coral cover, coral genetics, algal species distributions, biodiversity data (seagrass cover, fish, stingrays, dugongs, sharks, dolphins, turtles etc.), physical and chemical data (turbidity, salinity, sea surface
temperature, pH), nutrient data (Nitrogen, Phosphorus, Chlorophyll a, Sewage Plume Mapping), satellite imagery, 3D benthic data, graphs and publications. In the first instance, we will define common metadata standards and associated services for describing, publishing and discovering these highly heterogeneous datasets. There will be particular focus on automatically capturing the metadata necessary to support discovery, decision, and reuse. The discovery metadata will be made available to the ARDC. Future work will involve the development of more sophisticated data integration and access control mechanisms that support increasing levels of access to the metadata and data over time through - keyword search, ontological and spatio-temporal/mapping interfaces and RSS/Atom feeds.

10.2.2 Data Capture Project Descriptions

Australian National University

Earth Sciences: Development of a system to automate ingestion of data from seismology instruments, Data transformation and a starting point for the development of an online collection of Australian seismological data. High Level Software Functionality: features: streamline the current system for ingesting data and metadata from the instruments via SD cards, develop an automated system for creation of full specification SEED format files of the data and metadata from the instruments, and develop an automated system for the publication of RIF-CS metadata describing the datasets to Research Data Australia and the actual datasets to IRIS and other repositories as appropriate, within appropriate embargo restrictions.

Optical Astronomy: Acquire data from nightly Siding Spring Observatory Wide Field Spectrograph datasets and process for ingestion of data into the International Virtual Observatory Alliance and metadata into Research Data Australia. High Level Software Functionality features: acquire data from nightly WiFeS data sets and post process it for ingest of metadata into Research Data Australia and data into IVOA, automatically ingest data sets to both the ANU collection registry and IVOA and update RIF-CS collection, party and activity metadata, and service metadata as appropriate, to reflect dataset ingest.

Phenomics: Capture, management, security, distribution and publication of raw and analysed data from the Australian Phenomics Network and Australian Plant Phenomics Facility phenotyping platforms, and publication of metadata to the Atlas of Living Australia and Research Data Australia. High Level Software Functionality features: acquires data and metadata from the instruments, assembles and ingests individual datasets and accompanying metadata into PODD, and programmatically updates the collection descriptions held in PODD to reflect the ingest of additional datasets.

Humanities and allied disciplines: Amass large collections of digital material from humanities and social science researchers. High Level Software Functionality features: generates structured datasets and automatically extracted metadata using discipline appropriate schema, enables the user to augment automatically extracted metadata, and also build new collections, stores the data in the ANU Data Store and the metadata in the ANU Collections Registry, submits these files to discipline appropriate repositories and generates appropriate RIF-CS collection level metadata and submits to ANDS.
CSIRO

*Research Data Service: Multi Source (incl. Sensor Network) Data Capture:* The project will produce and deliver software technologies that will extend CSIRO’s newly established Research Data Service (RDS) system to provide generic configurable automated deposit of data and metadata from specific types of sources. This new technology will be validated by usage to capture data and metadata from CSIRO’s heterogeneous sensor networks and feed records to Research Data Australia.

Curtin University of Technology

*Curtin deployment and configuration of Institutional Metadata Repository and Research Data Portal:* Curtin University is currently developing research specific systems internally for the capture of strategically important research data, and streamlining of researcher workflows.

In order the make these projects more effective internally, and to contribute collected data into the Australian Research Data Commons, Curtin has identified two pieces of infrastructure to be developed under this ANDS funded project:

- A Research Data Portal to allow researchers to manage their data, object level metadata, and act as an endpoint point for data and metadata coming from the domain specific projects.
- A Metadata Hub to act as a collections level metadata store, to integrate with internal Party and Activity systems, and provide a harvesting endpoint for Research Data Australia.

Deakin University

*Enhancing Filtration Membrane Fouling Data Collection for Water Treatment Research:* Membrane filtration widely used by water/wastewater treatment and desalination industries. However, membrane fouling is one of the most severe performance limiting problems that has negative economic impact. The mechanisms governing membrane fouling is still not fully understood due to the complex nature of foulants and their interaction with membranes, especially in a mixed species environment. This project aims to develop image processing/analysis codes which will create new membrane fouling data from 3 dimensional images obtained by confocal laser scanning microscopy (CLSM). In addition, we will also develop tools to capture the instrument, processing and sample metadata. It is anticipated that the software developed and the collection of membrane fouling data and the accompanying metadata will significantly enhance the collaborations between researchers in membrane materials and water treatment at Deakin and other organizations to tackle the complex membrane fouling issues in water/wastewater and desalination and secure long---term quality water supply to Australian households and industries. The deliverables derived from this project will potentially be used and reused by many membrane materials scientists, water researchers and environmentalists around the world.

*Crystal Orientation Data Collection for Conversion to a General Data Type:* The Crystal Orientation Data Collection for Conversion to a General Data Type Project (DC20A) was undertaken by Deakin University 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 reuse 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.

The project established a standard format for each of the three types of diffraction data being generated from three different electron microscopes, then developed software to convert raw data into the standard format with appropriate labelling and transfer these files to appropriate storage. This software (CRYSTO) has been delivered to an open source repository and can be accessed at http://code.google.com/p/crystal-orientation-data-collection/. Also available on this website is the supporting documentation including user and technical manuals to enable installation and use of the software.

**Flinders University**

*Automated measurement of the responses of wildlife populations to climate change:* The aim of this project is to make the lizard data packages more broadly visible through the Australian Research Data Commons, and also publicly available through a Flinders University institutional research data repository. We intend to develop a system to convert CSV files to an appropriate format, and to generate and store RIF-CS collection, service, activity and party metadata. The RIF-CS metadata will reside within a metadata store that features an OAI-PMH harvest point and that will interface with other ANDS webservice (Identify My Data - PID; Register My Data; and the Digital Object Identifier webservice currently in development by ANDS). The data itself will reside in the data repository in an appropriate format.

**Griffith University**

*Smart Water:* Water end use study, data capture from instruments, reporting on domestic water use. To build a software system to collect data from remote smart water meters and integrate it into related data sets obtained from other sources and use metadata to describe the data used and the cross-relationships between data. The system will be used to build and integrate data sets to explore patterns of water use, the data sets, descriptions and mega-processes developed.

*Adult Stem Cell & Neurobiological Microscopy Instrumentation and Research Data Management:* To develop a software system to centralise the management of a large volume of microscopy image and related experimental metadata, allowing researchers within the National Centre for Adult Stem Cell Research ("NCASCR") to more effectively organise and analyse their biological imaging experiments.

Objectives of the project are to capture metadata from images generated from microscopy instruments, Import images in a standard format, allow web based browsing and management of image collections capabilities, share imaging collections, allow annotation of images and image collections, searching of metadata and it will generate and send "published" image collection metadata to Griffith’s Research Activity Hub.
James Cook University

Tropical Data Hub: This project will develop software to capture and create collection descriptions for a wide range of data sets out of the National Climate Change Adaptation Research Facility (NCCARF), Daintree Rainforest Observatory (DRO) facility - a high profile education and research centre, as well as the Wet Tropics World Heritage (WTWH) and the Australian Centre for Tropical Freshwater Research (ACTFR) centres.

Software will extract rich metadata from four years of environmental sensor data from NCCARF Biodiversity activity and DRO and real-time sensor data from next generation sensors as well as from systems of the WTWH and ACTFR centres.

The Tropical Data Hub (TDH) is a proposed 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.

These collections will be made available to researchers via the JCU Tropical Data Hub and will be harvestable via Australian National Data Service’s Research Data Australia System.

LaTrobe University

CMSS RLI Metadata Capture and Publication: The aim of this project is to leverage the existing ANDS services to support Australian researcher’s use of instrumentation at the La Trobe Centre for Materials and Surface Science facility (CMSS). This project will aid researchers in tracking their datasets and associated provenance metadata from creation at the instrument, in transfer back to institutional repositories, and right through to registration into the data commons. The capability that this project develops will also support collaborative data sharing through the use of persistent identifiers (IDs) that can track data location over time and across institutional boundaries. As a part of the NCRIS Australian National Fabrication Facility capability, La Trobe University's CMSS will also act as an exemplar facility in eResearch development and the use of ANDS services.

Macquarie University

Glycomics Repository: This is a two part project: Part I. The development of a world-class glycan reference repository (to be known as UniCarbDB). Part II. Metadata capture enabled from mass spectrometers being used to generate data to support glycan structure study. For more information about this work: UniCarb-DB: A Database Resource for Glycomic Discovery http://www.ncbi.nlm.nih.gov/pubmed/21398669

Papyri Data Capture: Metadata and digital images associated with papyrus manuscripts from Egypt will be captured to improve resource discovery, domestic and international access, and research collaboration. Parts of the same papyrus separated in the acquisition process can be recognised and digitally reunited; experts in particular types of texts can recognise and access items of interest to them; an important facet of the documentation for the ancient world can be explored online. Collection descriptions will be supplied to the Advanced Papyrological Information System, and transnational online initiative, and to Research Data Australia.
Monash University

Research Data Management of the Monash Weather & Climate Program (Climate and Weather): Monash researchers and research students create significant datasets, mainly through numerical climate model simulations. The proposed infrastructure will help to revisit and reuse the work carried out by research students and postdoctoral researchers who have left MW&C (Monash Weather & Climate) after finishing their degree or when their contract has ended. Particularly large amounts of data will be created in the coming years when the newly deployed Australian ACCESS climate model is evaluated by a team of postdoctoral researchers. These simulations will be done on the NCI computer platforms in collaboration with other Australian universities, the Bureau of Meteorology and CSIRO. It will be very beneficial for the MW&C researchers to be able to manage this data efficiently and publish it so that other participants will be able to process the data further. Funding of more than $2,000,000 has been secured for this work. In 2010-12 MW&C will carry out a suite of computer simulations to study the urban meteorology in a changing climate. Simulations will assess rainfall patterns in localised urban environments to inform the development, adoption, and operation of stormwater harvesting solutions in collaboration with the Institute for Sustainable Water Researchers, Monash University.

Biomedical Data Platform (Molecular Biology): MyTardis seamlessly facilitates the capture and annotation of protein crystallography data from the Australian Synchrotron MX beamlines, the deposition of the data to the researcher’s institutional data store, and the metadata publication across multiple research data repositories, including TARDIS (Monash-run), Protein Data Bank (worldwide) and the ARDC (nationwide). It has been integrated with Monash University’s High Performance Computing grid to process the raw data and store the output. Visit www.tardis.edu.au to view the public solution.

Tools for curating and publishing research data in the form of media collections (Multimedia Collections & ARROW): This solution, built for the Faculty of Arts, provides a workflow that enables members of the public to submit their life stories about adoption in text and sound files, subsequently able to be reviewed and curated before being published on the History of Adoption website and on the ARDC. The previous manual workflow, involving many systems, has been simplified significantly through the integration work of the project. http://arts.monash.edu.au/historyofadoption/

Capture and publication of Australian ecosystem data from a network of measurement sites (Ecosystem Measurements): This project seeks to directly address the lack of coordination in data collection, archiving and quality control across the active groups in Australia and to promote access to the integrated research data by both Australian and international researchers. These goals are closely aligned to the ANDS principles specified in the Final ARDC EIF Project Plan.

The infrastructure being proposed consists of the implementation of several software systems and associated documentation and training designed to:

- standardise and automate the collection, archiving and quality control of measurements from a network of instrumented towers;
- integrate complementary data streams from different sources into a single data and metadata repository;
• provide documentation and some training in the newly developed system; and
• facilitate the linking the data into a common research data space, through the Australian Research Data Commons

Capture and publication of data on the history of adoption (History of Adoption): The History of Adoption Project is an ARC-funded project undertaken by researchers from Monash and the Australian Catholic Universities. A central feature of the project is the collection of life stories as self-created sound files directly deposited by contributors, and the online publication of edited versions of these. The whole body of public data will be searchable, and collection records will be made available to ANDS for the Australian Research Data Commons. The project therefore requires infrastructure for data capture, enrichment and dissemination.

Data and metadata will be directly deposited by contributors: data as sound files both pre-recorded and directly streamed via Skype and Google Voice; metadata in the form of information identifying the source of the deposit and the historical context of the adoption described. Before deposits are accepted, depositors must sign off on ethical protocols determining their relations with the project and the project’s responsibilities to the depositors. Before data is made accessible to the public it will be edited for legal and ethical purposes, with additional metadata added at this stage. Different levels of access will be needed; in the first instance so that it is accessible only to its author, for review and confirmation; secondly, with full public access. At this point readers will be able to add comments to the data.

Data Publication to Interferome (MIMR/Interferome): Previously Ecosystem research data was difficult to access and hampered the actual research and analysis of the data.

The solution provided by the Ecosystem project has enhanced the research process and provided new research opportunities.

Overall, the solution cuts down the time significantly before the researcher can begin their research work and facilitates much simpler sharing of data. Due to these benefits the Ecosystem solution has been very well received by the OzFlux community of researchers. This adoption is a testament that the solution funded by ANDS is of significant value to the researchers in the community. The next key challenge is to enhance the Ecosystem solution to facilitate Ecosystem research on a larger scale at the national level.

This project realised a number of ground-breaking technological advances in the ecosystem distributed data system (Eddy). In particular, the project saw the development of an improved accessibility to the climate data and an auditable, searchable, distributed data system of research data. The Eddy system is fully integrated with the Research Data Australia Service so that the metadata of the research data is discoverable by the broader national and international research community. The Eddy system supports a range of scientific data formats, like NetCDF (Network Common Data Form), GRIB2 (GRidded Binary) which makes the creation, access and sharing of scientific data easy.

In addition, the Eddy system is a multiple-tier web-centric application and follows a proven classic design pattern, Model-View-Controller. Each component of the Eddy system is independent, so it is flexible for future development and can be easily maintained. The Eddy system also uses the Monash Large Research
Data Storage (LaRDS) service, which ensures that the research data reliably backed-up and secure. The solution can be viewed at: [http://ozflux.its.monash.edu/ecosystem/](http://ozflux.its.monash.edu/ecosystem/)

Queensland University of Technology

*Greenhouse Gas Emissions from Australian Soils:* Data and metadata capture from automated N2O gas sampling systems of emissions from Australian agricultural soils. Software written in eclipse will take the metadata as a "data package" and then sent to Metacat, a web-based server application with a database.

*Biodiversity:* Study the effects that humans have on the environment. Adopt acoustic sensing, database and web service technologies to capture audio data. Generate audio data and tags. Software to facilitate the mapping and transformation of data contained in the relational database, into a form that will be able to be ingested into QUT’s Metadata Hub.

*B150 BigJam:* A roster of 900 musicians improvising across a number of genres, performing and recording non-stop music for 150 hours. Software written using eclipse to facilitate the management of metadata relating to multimedia data captured. Management of metadata -- annotation, mapping and querying of multimedia data.

RMIT University

*Data Capture from High Performance Computing Multi-User Environments:* RMIT is currently a very large user of state and national high-performance computing (HPC) facilities. This project will develop and deploy software tools and applications which will be deployed for the NCI Supercomputer National Facility, the VPAC Supercomputer Facility and the RMIT HPC Facility.

The following simulation packages will be targeted:

- VASP (a Density Functional Theory code for simulation of material properties using plane waves)
- CRYSTAL (a Density Functional Theory code for simulation of material properties using Gaussian basis sets)
- SIESTA (a Linear Scaling (a Density Functional Theory code for simulation of material properties)
- GULP (Empirical potential code for simulation/modelling of material properties)
- DFTB (Density Functional-based Tight Binding)

University of Adelaide

*Automated capture and publishing of data generated on high throughput plant phenomic platforms:* Extract and curate data generated by the Plant Accelerator’s LemnaTec Smarthouse platforms. Publish the data to a navigable public repository, and metadata to Research Data Australia.

*Genomics Data Capture:* The University of Adelaide and SA Pathology joint Centre for Cancer Genomics will be characterising cancer genomes via whole genome and exome re-sequencing.

This work will be centred on a third generation gene sequencer. Each experimental run creates a set of sequences, associated settings descriptions, and analytic outputs along with experiment metadata. This set
is a collection that is described with metadata to support discovery, decision and reuse that is collected from experiment design/description, research group descriptions, information extracted from instrument settings and links to earlier work. The metadata collected will use relevant national and international standards, and be decided in conjunction with Bioplatforms Australia.

University of Melbourne

*Melbourne Neuropsychiatry Centre (MNC) Bioinformatics Development Project:* The MNC has one of the largest databases of brain scans and associated neuropsychiatric research data in the world. It has National and International collaborators using and contributing to the database. Broadly ANDS-related activities will involve:

- Building a workflow for automatic documentation of dataset segments used in individual studies and publications. This will include researchers, datasets, associated projects and publications.
- Building a workflow for automating creation of citable persistent identifiers for unique studies and linking with publications.
- Building a software to automate capture of public facing metadata to University of Melbourne Registry which will deliver collections metadata to the ARDC.
- MNC has 270+ publications resulting from datasets stored in the MNC database.

*Youth Research Centre's Life Patterns Project:* Longitudinal qualitative and quantitative survey data capture and reuse: "The Youth Research Centre's Life Patterns Research Program maintains an extensive qualitative and quantitative data base on a cohort of 2000 young Australians who left secondary school in 1991 and of a second cohort of 3000 who left school in 2005. With ARC funding through to 2014 for annual quantitative and qualitative data capture for the second cohort (Gen Y) and biannual data capture for the first cohort (Gen X), this activity aims to enable wider access and use of the data by developing the infrastructure to make sets of the existing data available for reuse, streamline capture of new data so that it is more readily available for reuse, and build the capacity to efficiently respond to future requests for derived data sets.

Appropriate structures for the capture of relevant metadata (compliant with DDI2, DDI3 and RIF-CS schemas) and tools to extract this metadata from workflows will be developed.

*Video data in the Social Sciences. Optimising Metadata Capture, Data Sharing Procedures and Long-term Reuse:* The University of Melbourne has an especially rich humanities and social science research community that utilises video as its primary form of data capture. The increasing use of video as a research tool poses particular challenges for aggregated data storage initiatives. This project will integrate metadata capture facilities at selected sites within the University of Melbourne as part of facilitating sharing and reuse. The project will address current metadata issues associated with large-scale audio-visual repositories and workflows to enable efficient generation of metadata, ensuring that stored video data is accessible and searchable through the ARDC. The project will:

- Develop software to automate the capture of metadata from existing mature video storage systems developed by the ICCR (International Centre for Classroom Research),

30 September 2012

[ands.org.au](http://ands.org.au)
Develop and where possible - utilise existing infrastructure to identify generic workflow tools that will enable rich knowledge of data sets, access services and parties to the research to be systematically (RIF-CS) captured from the researchers,

- Develop standards compliant video data and metadata deposit services.

These are generic goals which are broadly applicable to activities elsewhere within the university, for example in the Faculty of Architecture, Building and Planning and the Faculty of the VCA and Music.

Federated Neuroimaging Collections in the National Data Commons: DaRIS is a raw data management system based on the Mediaflux digital asset management platform and has been in operation for the last 3 years at the Neuroimaging Computational and Data Management Facility (CDMF). There it has been used to routinely receive MR images from researchers and organise them into a subject-centric data model, ready for access by project members. It hosts over 70 mouse and human projects, each with many tens of subjects and some with time-dependent data.

- Map DaRIS project-metadata to the ANDS schema
- Write a DaRIS service to populate ANDS-compliant metadata,
- Develop an adapter to harvest the ANDS-compliant metadata from DaRIS
- Connect identifiers within DaRIS to ANDS persistent identifiers (PIDs).

Humanities and Social Science Research Data at the University of Melbourne: The University of Melbourne has one of the most rich and diverse humanities and social science (HASS) research communities in Australia and is well ranked internationally. HASS researchers at Melbourne generate and hold valuable data sets and associated materials that are currently not easily discoverable, accessible or configured for further research purposes. This project will build infrastructure (tools and services) to connect this diverse community with the UoM Registry (Vitro) which will in turn communicate the relevant metadata to the ARDC. The project will:

- Develop and utilize existing (OHRM-based) infrastructure to identify generic workflow tools that will enable rich knowledge of data sets and related materials, access services and parties to the research to be systematically (RIF-CS) captured from the researchers.
- Development of a generic web services-based data capture tool to be used both by researcher staff, data librarians or other staff in the data management fabric. This will be based on the 'pre-register' work done for the Australian Women's Register in 2009
- Develop standards compliant 'access service' descriptions
- Ensure project, data, party and service descriptions concord with Data Documentation Initiative (v2&3) requirements.
- It will inform the development and utilisation of digital and analogue archival preservation, curation and access systems for the University

Capture of Complex Data to Support Clinical Research in Cardiovascular and Neurological Medicine: Complex physiological data is routinely collected on patients as part of clinical care (echocardiography, intravascular ultrasound, x-ray angiography, optical computerised tomography, patient clinical data, etc.). However, this rich multi-model data is not usually subjected to subsequent analysis nor is it made available to researchers
from other disciplines for novel analysis. Making this multi-model data available along with patient outcomes such as morbidities will provide the opportunity for collaborative groups to employ novel strategies to developed assessments and models based on this data. This project will form necessary base of making multi-model data collections available, enabling the establishment of new links between biomedical research groups in engineering, physics and bioinformatics. This project will occur in collaboration with BioGrid Australia where it will use the access, de-identification and privacy protection protocols already established there.

Founders and Survivors Project: "The Founders and Survivors Project (http://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 19th century and their descendants to create a population database of national and international significance for historical, demographic and population health researchers. This project will:

- Develop a toolkit based around the projects XML/TEI workflow for further relevant records sets to be systematically ingested into the population database,
- Build the infrastructure to enable persistent identification and descriptions of derived data sets produced on request from the population database to be made available to the ARDC

University of New South Wales

ARDC Linked International Glycomics Repository & Instrument Data Capture: Data and metadata capture from mass spectrometers, hosting of world-class glycan repository.

An international antibiotic-resistance gene cassette database: Software to allow researchers to contribute and find data on grammatically-defined bacterial gene cassettes

Data capture and integration across multiple platforms: Manage and integrate data from the Analytical Centre, the Membrane Library, and the Electronic Lab Notebook

ANZNN Neonatal Data Capture Portal: The primary objective of the project is to provide a web service interface for ANZNN to collect and manage its data from all neonatal units participating in ANZNN activities. Through this project, ANZNN aims to achieve the following objectives:

- To provide a secure web service interface for NICUs and NUs designated staff to review the minimum dataset requirement, define the meta-data mappings, upload dataset files and have the web service interface run verification processes on the loaded datasets;
- To set up verification software for data based on a pre-defined and agreed set of data and meta-data rules;
- To provide feedback through the web service interface to NICUs designated staff regarding their uploaded dataset files;

30 September 2012
• To provide a secure data entry form for NUs without patient record management system to create full data records from their written clinical notes; and
• To establish ANZNN data commons for research by other research communities, for example health economics professional and disability children education researchers, through contribution to Research Data Australia.

Through the immediate feedback built in the web service interface and provision of minimum dataset requirement, the project will assist the participating NUs to map their own data collection to ANZNN standards correctly, therefore, to improve the process of ANZNN data integration. The standardised data collection can then be used for comparison research and benchmarking of the care provided by NICUs and NUs.

University of Newcastle

Data Capture for the Data Commons: The Health Behaviour Research Group (HBRG) is a key member of the Priority Research Centre for Health Behaviour at the University of Newcastle. HBRG enjoys a strong international reputation for quality research in behavioural and public health approaches to health promotion, health service evaluation and cancer control.

Collection and management of survey data is a fundamental activity for HBRG. On commencement of a research project, HBRG researchers determine characteristics of the proposed survey, for example characteristics of the (subject and control) participants who will take part in the survey; the questions to be asked; and the statistical analysis to be performed. The survey instrument, once designed, is then administered using paper-based questionnaires, or more recently using purpose-built computer software. Once datasets have been collected, the group’s researchers analyse it, leading to suggestions for change to the processes and procedures performed within the health care system.

University of Queensland

Spatially Integrated Social Science: Capture of derived spatially-integrated data from interrogation of data contained in the social science data archive that includes both legacy databases (e.g., Australian Bureau of Statistics (ABS) and Australian Electoral Commission (AEC) data) as well as derived datasets and newly generated data.

Microscopy/Microanalysis Image and Data Repository: It will enable images and metadata to be captured/extracted directly from the instruments to a users’ personal workspace - but then support selective upload of images/data by owners to a secure CMM image/data repository;

• This project will build on existing infrastructure established during the GRANI project which in turn uses international standards such as the Open Microscopy Environment (OME) metadata schema
• It will focus initially on the Tecnai F30 TEM

DIMER Diffraction Image Repository: Project to capture and share X-data and images generated from the Rigaku X-ray diffractometers in the UQ Remote Operation Crystallization and X-Ray Diffraction Facility
Linking the EMBL Australia EBI Mirror with the Australian Research Data Commons: The Aims and Objectives of this project are:

Component A: to populate Research Data Australia (RDA) with collection descriptions of data held in the EBI databanks.

This will specify and implement automated systems to identify bio-molecular data collections already held in the EBI databanks that are "Australian-associated", extract metadata from the EBI databases for these data, generate RIF-CS collection-level descriptions, make these collections discoverable through RDA, and enable navigation from RDA to the corresponding primary data entries or sections in the EBI Mirror. "Australian-associated" data in this context is relatively broad and may be defined as: Data submitted from Australian-based researchers; Data associated with sets (and subsets thereof) of Australian species (e.g. native, introduced, agricultural); datasets of relevance to Australian-based researchers (e.g. related to common diseases facing the Australian population). The overall approach will enable researchers in a wide range of fields that are not necessarily bio-molecular domain experts (e.g. marine science, ecology, climate studies, agricultural science, health science etc.) to use RDA to easily locate and use relevant molecular data in their studies.

Component B: to enable submission of descriptions to RDA for data associated with secondary analyses performed using the Australian EBI mirror.

This will specify and implement automated systems to extract metadata from secondary analysis of data (either EBI data or user-uploaded data) performed by researchers using the NCI-SF in Bioinformatics instance of Bioflow, to generate appropriate RIF-CS collection and service descriptions for datasets generated through the secondary analysis, make these metadata discoverable within the RDA, and enable navigation from RDA to the corresponding Bioflow workflow and EBI data in the EBI mirror. This will enable complex
bioinformatics workflows that can be applied in the re-analysis of molecular data, to be found by a wide audience and reused.

Component C: to deliver collection descriptions for EBI data aligned with BioPlatforms Australia (BPA) themes to RDA.

This will engage BPA and other relevant R&D communities to help conceptualise and specify views of Australian EBI Mirror data within the context of BPA theme projects, and organise data within the Mirror to constitute these views and make them discoverable through RDA. This will enable the data produced through the NCRIS/Super Science BioPlatforms Australia investments to be presented through the EBI mirror and RDA.

University of South Australia

*Development and testing of a data capture tool for instruments at the Ian Wark Research Institute:* The project’s aim is to develop a MetaData Capture Tool, a semi-automatic tool to facilitate capture of both data and metadata from the IWRI TRIFT V and Mastersizer 2000 instruments.

University of Sydney

*SKAMP Data Capture:* SKAMP, or SKA Molonglo Prototype, is a joint University of Sydney and CSIRO project, based on the Molonglo radio telescope. This project will automate making the SKAMP data available and allow timely follow-up of celestial events (such as supernova explosions). The key objectives are to enable observation of targets-of-opportunity, to drive robotic telescopes, to alert researchers to search and review archive data in areas where transient objects have occurred, and to alert the international community to events that are detected with SKAMP.

*NSW TARDIS Node:* The project will establish a single repository for macromolecular protein crystallography data in NSW - the NSW TARDIS Node. This is a discipline-based as well as an institutional initiative. There are four instruments in NSW and the ACT producing data suitable for inclusion in TARDIS. These instruments are at University of Sydney (Biochemistry), University of New South Wales (Physics), the Victor Chang Cardiac Research Institute and the Australian National University.

*AMMRF Live Cell Microscope Data Capture:* Experiments conducted using high-end microscopy instruments currently record information in multiple dimensions, such as 3D and 4D, often with multiple detectors. Data collections are therefore increasing in size as microscopy techniques evolve, often resulting in extremely large images that are difficult to manage using current systems. Images also often require post-acquisition processing steps where the data moves between analytical platforms, which can be situated in different sites.

These issues have created data management requirements for a growing community of microscopists, necessitating development of infrastructure to transfer experimental data from instruments to a data repository. Automated capture and storage of associated metadata is also critical, in order to enable data to be discoverable and accessible by the data producer in addition to other researchers. An existing Neat AMMRF PfC project being undertaken by Intersect is developing a data management and ingest system for
microscopy data from MicroCT and AtomProbe scanners. The repository supporting this system is based at the University of Sydney. The information collected is being made available to all AMMRF researchers and other researchers who have access permissions. The Sydney microscopy data capture project will complement the AMMRF NeAT project by catering for additional instruments, extending the scope for the repository that project is developing.

Data capture from the Olympus systems, to be added to the development of the data management system, will enable collection metadata to be discoverable in RDA if a researcher so desires. Researchers will benefit from improved data management capability of their own data, sharing of data and data reuse.

**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.

**FieldHelper (ExCite9):** The project will develop a cross-platform user-friendly and flexible software tool with standards compliant workflows for managing the flow of research data from fieldwork-based research projects through repositories into the Australian Research Data Commons. Discoverability and reusability of such data will be improved through facilitating standards compliance at the time of data capture, and interoperation with other ANDS services and projects. The tool will be designed to be adaptable to diverse areas of study in the humanities and social science domains.

**AgDataCapt: Capturing Agricultural Data:** Standards-based body of data and metadata will be coordinated and include data from sensors on soil, water and rain, greenhouse gas and carbon data, and weather. Generic and extensible tools will be developed to integrate appropriate standards across the areas of data capture. The areas of data capture are: Soil moisture, soil and air temperature, radiation, 3D wind speed and directions, CO2, water vapour, weather and tree water use data, greenhouse gas emissions from soils; GPS-referenced data on wheat and barley crop inputs and performance; Soil data from sensing system for monitoring of agro-ecosystems for sustainable landscape; Soil data and spatial prediction functions for soil variables; and Data from farm-based private rain gauges.

**University of Tasmania**

**Redmap Australia:** 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.

**Data Capture of state-wide hydrological datasets:** The aim of the project is to capture and publish data from the CSIRO and Forestry Tasmania sensor webs. The data will be exposed by two sensor observation services (SOS) serving twenty two sensors, and a THREDDS server providing forecast data.
University of Technology, Sydney

Maximising the Benefit 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 reuse 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.

University of Western Australia

Deployment and Configuration of Institutional Metadata Repository: The overall aim of the UWA Data Capture projects is to ensure that the University of Western Australia has a metadata store for capturing metadata about research data collections, and that the University is in a position to contribute metadata in the RIF-CS format to the ARDC. The projects will also implement data capture tools and systems for three strategic research areas: rock art research, video data for marine ecology research, and integrated data capture for characterization and analysis in bio-imaging.

The UWA1 project specifically aims to ensure that the University of Western Australia has an institutional metadata store capable of capturing metadata about research collections, and that metadata from this store can be contributed in the RIF-CS format to the ARDC.

UWA Rock Art Studies Data Management: The desired result of the Rock Art project is to improve the ability of UWA archaeologists to maintain their data for personal use, and to share it with collaborating researchers and with the wider research community when appropriate. Making the data available to indigenous communities is also a goal.

The scope of the project is currently to provide a solution that works specifically to meet the needs of the Rock Art group within the UWA Archaeology department. There is however a strong desire to develop the system so that it may be used for other similar groups within Australia and for other archaeology groups.

The Rock Art Data System (RADS) solution will be connected to the UWA VIVO metadata hub, which in turn will be connected to the Research Data Australia metadata hub.

Organisation of Australian Underwater Video and Still Imagery: As technology for capturing information about marine environments improves and becomes more widespread, the volume of data collected has been increasing by several orders of magnitude.

Marine researchers in Australia and overseas are using ever more sophisticated imaging technology which is producing valuable data.

However the cost of creating this data can be significant and thus it is imperative that its value is fully extracted by sharing it with the wider research community to prevent overlap and find new results that aren’t necessarily obvious from smaller data sets.
The need is to create a system for managing the raw images, data that describes the various marine objects and properties (e.g. fish length, fish species, water depth, water temperature, salinity). It should also maintain a record of metadata that describes the research study being conducted, provides information about the relevant grants and organisations and importantly the researchers involved in carrying out the study.

*Integrated data capture for characterization and analysis:* The desired result from the CMCA project is to implement software and hardware infrastructure that allows the microscopy data (raw image and spectral data) to be stored, and searched based on appropriate metadata. This metadata includes machine settings and other directly available data. It also includes data that cannot be directly captured from the instrument such as information about the researcher, research group, grants funding the research and sample information (e.g. mouse brain). This latter type of information currently requires much manual effort to collate and thus automating its collation with the raw data is of significant benefit to CMCA management.

**University of Wollongong**

*Biomechanics Data Capture Project System:* Data and metadata capture from multiple instruments in biomechanics lab. In more detail: Kinematic data; OptoTRAK 3020 system, Force data; Kistler Multichannel system, Electromyography (EMG) data; Noraxon Telemyo system, Ligament laxity; Dynamic Cruciate Tester, Pressure data; Novel pressure measurement system, Surveydat; Survey Monkey

*Remote Sensing Spectral Library:* This project will build a system to allow the Australian remote sensing community to collate, share and discover new and existing spectral libraries. In particular it will provide recorded consistent metadata and a consistent method for publishing, discovering and assessing this information.

Analogous to this project is the AusCover component of the Terrestrial Ecosystem Research Network, which is providing a system for the publishing and discovery of remote sensing satellite data, and developing standardised protocols for the collection of spectral data for remote sensing. This project will be a natural extension of work already happening nationally, and provide the missing link not covered by the AusCover facility.

### 10.2.3 Metadata Stores Project Descriptions

The funded Metadata Stores projects are all based on the following list of expectations and deliverables. Optional deliverables vary from project to project.

A metadata store deployed at institutional level should fulfil the following requirements:

- **R1:** To manage metadata about data collections held at the institution
- **R2:** To enable discovery and reuse of data collections held at the institution
- **R3:** To support strategic planning for research in the institution
- **R4:** To ensure high quality of metadata to support the other requirements on a metadata store

A metadata store deployed at institutional level can have the following functions:
F1: To aggregate metadata for data collections managed by subunits of the institution, where the subunits are already managing such metadata (supports R1)

To expose metadata for data collections managed at the institution, in a variety of disciplines, to external discovery (supports R2)
  - F2: At an institutional level (e.g. research profiles, researcher profiles)
  - F3: At a national level; not limited to ANDS (RDA)
  - F4: At a discipline level

F5: To expose and manage metadata about data collections within the institution at the object level, coupled with the exposure and management at the collection level (supports R2)
  - Including data for reuse (reproducing the collection), data for discovery (granular to the individual object)
  - (Note: ANDS does not require object-level management to be provisioned by the one system across the university, or necessarily by the same system as for collection-level management)

To impose consistency and coherence over metadata descriptions, including metadata descriptions aggregated from various subunits of the institution (supports R4)
  - (Preferably by following metadata standards appropriate to management of research data; RIF-CS is not such a standard)
  - F6: Including uniform use of vocabularies and schemas
  - F7: Including aligning metadata descriptions with institutional sources of truth --- such as HR (persons), Research Office (grants)
  - F8: Including aligning metadata descriptions with national sources of truth --- such as NLA (persons), ARC/NHRRC project registry (grants)

F9: To register new data collections as they are created and come to be managed within the institution (supports R1)

F10: To register and manage core attributes of data collections required for effective data management (supports R1)
  - Including managing access (embargo), current location (online or offline --- loose coupling with storage), rights, retention, provenance, citation, audit, proprietary tools & formats, policy framework (data management plan, ethics)

F11: To generate analytics, network analysis, researcher profiling, and other internal reports on the extent and breadth of collections described in the store (supports R3)

To realise these objectives, the following deliverables are expected out of the metadata stores new funding:

Mandatory:
  - D1: A working feed of records describing collections and associated activities, parties and services to Research Data Australia, in the current version of RIF-CS (1.3), demonstrated to meet the quality requirements for RIF-CS records set by ANDS. (realises F3) (realises F6)
  - D2: Coverage in the Research Data Australia records feed of collections from at least three distinct faculties (or equivalent organisational units) within the institution (realises F1)
• D3: Demonstrated alignment of metadata records about parties with an institutional name authority (HR or Library), with the authoritative form of the name sourced external to the metadata store, and with new researcher descriptions added to the metadata through regular update from the name authority (realises F7)

• D4: Demonstrated alignment of metadata records about parties with the ARDC Party Infrastructure Project, with researcher descriptions contributed to the NLA, and with People Australia identifiers for researchers recorded against researchers (realises F8)

• D5: Demonstrated alignment of metadata records about activities with institutional and external sources of truth (Research Office, ARC and NHMRC grant registries), with the authoritative description of the activity sourced external to the metadata store, and with new researcher project added to the metadata through regular update from the sources of truth (realises F7, F8)

• D6: Demonstrated workflow for registering new collections in the university; this can include automated update, or semi-automated (notification-based) (realises F9)

• D7: Demonstrated ability to manage the following aspects of the collection lifecycle through recording and exposing relevant metadata: (realises F10)
  • embargo dates for collections, where applicable
  • current online location of collection (on internal store or external store)
  • current offline location of collection
  • intellectual property rights (licensing, restrictions on reuse)
  • retention policy (disposal date, deposit date)
  • policy framework (data management plan relevant, ethics forms relevant)

Optional:

• D8: A public researcher or research profile portal, exposing publishable metadata about the research data being held at the institution. (realises F2)

• D9: Demonstrated ability to feed a selected subset of the collection records relating to a particular discipline to a discipline registry, following the metadata schema and conventions of that registry (realises F4)

• D10: Demonstrated ability to manage the following aspects of the collection lifecycle through recording and exposing relevant metadata: (realises F10)
  • citation requirements (authoritative identifiers, including DOI, preferred citation format)
  • citation tracking of collections
  • audit information (refer to publications audit)
  • proprietary tools and formats used in collecting the collection

• D11: Strategic reporting on contents and coverage of metadata store for internal use (realises F11)

• D12: Storage and exposure for discovery of object level metadata, and alignment of object level metadata with collection metadata (i.e. ability to navigate from object metadata to collection metadata; update of object metadata aligned with update of collection metadata) (realises F5)

• D13: Storage and management of technical metadata for object and collection reuse, including software and equipment descriptions, methodology, and data interpretation (realises F5)

Procedural Deliverables:
- **D14:** Project Management Plan, using ANDS template, specifying the details of the planned activity, with risks, schedules, etc.
- **D15:** Quarterly Progress Reports, using ANDS templates
- **D16:** Final Report, using ANDS templates
- **D17:** Deposit of any software (including style sheets and schemata) developed in the project for achieving other deliverables, and that can be usefully used outside the institution, in either Google Code or SourceForge, including:
  - a Google code comment and tag or SourceForge summary and tag containing the text "ANDS-funded"
  - Developer manuals where applicable, to facilitate reuse
  - Deployment manuals to facilitate external deployment
  - User manuals to facilitate use
- **D18:** A source code report, if any software is developed and publicly deposited under D17
- **D19:** A User Acceptance Test online survey

### 10.2.4 Public Sector Data Project Descriptions

**AuScope**

This project builds on the work undertaken in the SISS Project by:
- deploying the SISS offering at Government data providers with data holdings important to national research priorities and CSIRO flagships
- developing some of the component services and functional capabilities needed to realise a spatial information data commons within Australia supporting linkages with academia, research, public and private sector data and service providers and users
- leveraging the Open Geospatial Consortium (OGC) and ISO standards that have been adopted by a number of capability areas and institutions in Australia and worldwide for which various reference implementations already exist for information services and portal interfaces
- combining and building on existing expertise in this foundation area in order to assist multiple capability areas and institutions that have interest in geospatial web service middleware and clients.

**CSIRO WRDM**

The CSIRO Water Resource Data Management (drawing from the Water for a Healthy Country Flagship Sustainable Yields) project is developing tools required to automate the capture of data collections, develop metadata for harvesting into the Australian Research Data Commons (ARDC) to increase their discoverability and potential for reuse. It will include:
- Development of data translation toolsets and implementation of metadata standards.
- Establishment of infrastructure to allow direct access to datasets allowing researchers to either upload subsets of data direct to their desktop or copy complete datasets to appropriate advanced scientific compute facilities.
Establishment of a Water Research community data registry/repository building on the capabilities developed in both CSIRO Marine & Atmospheric Research/TPAC and AuScope initiatives.

Establishment of middleware systems that track and inform researchers about the licensing arrangements for each of the datasets CSIRO holds.

Population of the CSIRO repository with metadata and population of metadata in to Research Data Australia and the Water Resources Observation Network.

It is anticipated that other datasets will be added on completion of the project and establishment of the system.

**AODN**

Core data captured routinely from research vessels managed by CSIRO and AAD will be published in near real-time in a coordinated way that enables ready access to, and combination of the datasets via the Australian Ocean Data Network (AODN). These vessels are major national marine science facilities and are the source of a significant proportion of Australia’s in-situ blue water ocean research data. The AODN, when adequately populated, will be an online network of marine and coastal data resources, which will include data from the six AODC Joint Facility (AODC JF) partner agencies and other data providers, supported by standards-based metadata, and will serve data to support Australia’s science, education, environmental management and policy needs: Australia’s digital ocean commons. Publication via the AODN will provide the public with a simple access point for these core vessel-sourced data and ultimately other vessel datasets not currently routinely published. As well as being valuable data in their own right, the work associated with standardising/automating the publication of the routinely captured instrument data, could act as an exemplar for real-time data publishing in other domains. Vessel managers will also work with ANDS to automate the provision of researcher and project information relating to use of the ships to RDA. Collection level metadata would flow to the RDA along with the deep metadata, via the AODN. The AODN’s data turbine middleware will be harnessed to visualise the core instrument-sourced data.

**AustLII**

This project develops software to include information about the 300+ free access Australasian legal databases located on the Australasian Legal Information Institute (AustLII) system in the ANDS discovery services. It also develops software and automation procedures which will be tested and used, during the project timeframe, to make at least 40 additional ‘public data collections’ that are important for legal research available (via AustLII) and more accessible to researchers, so as to ensure greater use and re-use of these existing data resources. Outcomes will be greater awareness of accessible legal data, particularly in disciplines outside law, and improved research in a broad range of different research communities in Australia and overseas.

**Powerhouse Museum**

The primary aim of this project is to establish and populate the Cultural and Historical Collections Metadata Exchange with Collection Level Descriptions conforming to the RIF-CS schema for delivery into the Australian
Research Data Commons (ARDC). This includes the establishment of data capture processes, data standards and infrastructure amongst museum data partners in the project and the establishment of protocols to enable automated transfer of the collection descriptions into the ARDC. The Exchange will be hosted by the Powerhouse Museum, drawing on established infrastructure and in-house expertise in terminology development. A number of additional services and tools are also proposed to extend and enhance the data set for researchers and to establish automated harvesting arrangements. The project will work closely with HASS researcher communities, primarily through its research partners including Monash University, Flinders University, RMIT University, University of Sydney and through the Australian Academy of the Humanities, and the Council for the Humanities, Arts and Social Sciences.

The principal contributors to the metadata repository will be the major museums which are affiliated with the Council of Australasian Museum Directors. These museums hold some of the largest and most significant museum collections of interest to HASS researchers. Museums Australia will assist in gaining participation of regional and local museums.

**GeoScience Australia/OSDM**

Geoscience Australia and the Office of Spatial Data Management have three catalogues which hold metadata of significant value to researchers. ANDS will analyse the catalogues and endeavour to map to RIF-CS.

Activities include:

- Analyse ASDD, Geocat and Geomet catalogues and if feasible, map all or part of catalogue records to ISO19115;
- Analyse any residual records which did not map to ISO19115 and map to RIF-CS where feasible;
- Crosswalk ISO19115 records to RIF-CS format;
- Install OAI-PMH harvest point at GA;

In order to harvest the resulting RIF-CS feeds from Geoscience Australia to Research Data Australia.

### 10.2.5 Applications Project Descriptions – Projects commenced as of 30 June 2012

**Tropical Data Hub and Tools (AP01/AP02)**

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 will create 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 will in essence be a synthetic data product to solve specific scientific and socio-economic questions related to land use and conservation in the tropics. This project will be an exemplar of the power of the ANDS data centric research model and also demonstrate the value of tropical data integration to research and policy/planning. Target product consumers will be
researchers interested in terrestrial ecology and climate change, local and state governments and agencies as well as land and real estate developers. We are unaware of any similar tool set based on rigorous observational and computational data.

Marine Video (AP05)
This project will provide funding to support the development of a database framework applicable to marine imagery that will facilitate the timely delivery of standardized, quantitative estimates of ecologically relevant indicators (such as absence/presence, percent cover, abundance and distribution of benthic organisms and associated substrates) through the analysis of visual data being produced by AUV, BRUVS, ROV and UTV systems. This will be accomplished through the funding of two software engineers to develop the database and access tools to allow the extensive datasets currently available at our collaborating institutions to be consolidated. An end user workshop will be held to solicit input from our end user community to establish a set of requirements and specifications for the analysis tools and annotation framework.

Validation of genomes and transcriptomes with proteomic data (AP11)
High throughout (“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, and the tool being developed in this project will greatly facilitate these analyses. The project outcomes will be 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)
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. This project aims to: 1) integrate heterogeneous data sets from POS data layers with other geo-spatial build environment information and demographic and socio-economic status data, 2) to develop a web based geospatial data analytic and visualisation tool that would support two new opportunities: 1) In research: it would enable 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. 2) For urban planning practice: it would enable 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)

The SMART Infrastructure Facility will develop a ‘multi-utility dashboard’ that will offer infrastructure analytics based on data provided by public agencies and private operators. The online dashboard will allow analysts to develop new insights into spatial, technical, social and economic issues associated with regional and urban infrastructure development. The multi-utility dashboard will act 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 will constitute a robust foundation for powerful infrastructure analytics. The multi-utility dashboard will offer a flexible collaborative platform to researchers, business analysts and local planners. This project will contribute to enhancing Australia’s innovation capacity.

Soils to Satellites (AP15)

The project will allow for the identification, exploration and comparison of similarities and differences across various environmental sites across Australia. A tool will be developed to integrate plant and animal species occurrence/community data; plant and soil genetics data; and vegetation cover data (sourced from The Atlas of Living Australia, TERN/TREND, AusCover and Bioplatforms Australia) from 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 these data can indicate the relatedness of the communities and turn over in community composition along environmental gradients. Differences between these relationships and the spatial arrangement of the sites may indicate environmental drivers. 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 (AP16)

This project will build a system for the combination and subsequent distributed web based viewing that allows the inspection, annotation and analysis of multi-modal imaging data (MRI, classical histology) at multiple scales from multiple sources. This will be applied initially as a tool for neuroscience research to investigate the mouse brain, where a major challenge associated with the combined viewing of this data is that it simply is not possible to display the full resolution of all the data modalities in any one plane with current display hardware. The system will be of great use to the anatomy/neuroscience research community – both in the creation of an interactive mouse brain atlas, which can be used for both teaching purposes, and as the benchmark for comparison of wild-type to non-wild-type (e.g. disease) specimens. The tool to be developed here will greatly facilitate the identification, evaluation and quantification of the physical manifestations in both naturally occurring, and experimentally-induced genetic diseases in mice which are often used as models for studying human genetic disease.
Multimodal kidney imaging data integration (AP19)
This project will develop a tool that will inform future potential non-invasive, imaging based procedures for monitoring kidney disease progression. The project will build 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 on living subjects). The work will be carried out using data collected from rat specimens and this project will result 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.

Brain and Mind Research Institute data integrator (AP24)
Mental health and Neuroscience research programs associated with the University of Sydney (Youth Mental Health, and the Healthy Brain Ageing program) are undertaken at the Brain and Mind Research Institute. A number of these research studies involve health condition improvement programs (cognitive training, exercise, art, etc.). Research subjects from both studies are routinely scanned in a MRI scanner and undergo a standard battery of neuropsychological and electrophysiological testing. This data is enriched by contextual information critical for use and reuse. Research subjects themselves are contextualised by clinical data of essential value to the researcher. Although existing tools support characterisation, management and sharing of scans and test results, there remains an important and unmet need for tools enabling the application and management of persistent links relating research data, contextual data and clinical data (de-identified in the case of human subjects) and integration of these managed data types which would enable new types of statistical analyses to be run across the combined research scans and clinical records (such as the case of comparing a research brain scan and the results of a clinical cognitive test). Integration will also be of direct value to clinicians caring for patients who are also research subjects. Research findings would be returned to referring clinicians, providing additional diagnostic support.

Climate Change Adaptation Information Hub (AP25)
This project aims to build a software system that acts as a central information hub for researchers in the Climate Change Adaptation research domain. The system will be built to allow users to deposit research data with associated metadata descriptions into a central managed storage infrastructure. The system will also enable a variety of search types (including spatial and temporal searching) of metadata in addition to other discovery tools to locate relevant research data stored in the information hub. The benefits of the Climate Change Adaptation Information Hub include 1) collecting all associated research data relevant to climate change adaptation into one managed and searchable location, 2) enabling more effective sharing and collaboration with colleagues.

Marine Virtual Laboratory Information System (AP26)
The Marine Virtual Laboratory Information System (MARVLIS) will build on, and enhance, the software infrastructure developed by the NeCTAR early activity Marine Virtual Laboratory (MARVL), a system designed to integrate observations and model simulations for any user-defined region and time period.
Using the Derwent Estuary in Tasmania as a pilot study area MARVLIS will develop software to utilise outputs of the MARVL and demonstrate the value of these products in aquaculture and fisheries management and to Tasmanian State Government environmental assessment programs. Integration of MARVLIS software into MARVL ensures the portability of the software tools to any region.

**Cancer Genomics Linkage Project (AP27)**

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 reuse of these DNA/RNA datasets of international importance by research biologists and clinicians is currently limited by their ability to access and use computational and data infrastructure. This project will develop software that will enable researchers and clinicians to access, integrate and transform unique and valuable cancer genomics datasets from the ICGC and other sources and to analyse them using the power of the closely aligned NeCTAR-sponsored Genomics Virtual Laboratory software system.

**Primary production in space and time (AP28)**

The project will fuse disconnected data sources—weather, remotely sensed land-surface observations, CO₂ and water flux measurements, hydrograph data and remotely sensed CO₂ concentrations—to generate a time-varying field of gross primary production (GPP, the most fundamental function of all ecosystems) across the Australian continent. It will develop 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 will be of strategic value to climate and carbon policy makers in Australia, and of great utility for benchmarking ecosystem models in Earth system science.

**Species Distribution Records (AP03/AP30)**

There is currently a scarcity of transparent online tools which integrate species distribution data, locality data with climate change scenarios in an integrated fashion which will facilitate the modelling of current and future species distributions based on climate scenarios. This project will collaborate closely with James Cook University (JCU) to develop a tool that reuses data available through the ALA and the JCU Tropical Data Hub to explore the potential impacts of climate change on a wide range of species in Australia; and engage in improving our understanding of the species and the modelling of species distributions.

**AURIN & ANDS – North West Metropolitan Region of Melbourne Data Access, Integration and Interrogation and Demonstrator Projects (AP31)**

The purpose of this project is to facilitate access to a myriad of data sets for the Melbourne North West corridor. This data test bed will provide access to a number of datasets to researchers via the AURIN portal. Such 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 project is to be conducted through the Centre for Spatial Data Infrastructure and Land Administration (CSDILA) at the University of Melbourne and involves three faculties across the University of Melbourne. The aim of the project is 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 will be demonstrated through four demonstrator projects which cover four of the most pressing issues facing the North West Melbourne Region they are: built environment and health, housing affordability, economic productivity, and transport and sustainability.

**Human Proteome Browser (AP32)**

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 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” being developed in this project will integrate a wealth of currently publicly accessible, yet disparate protein data and function as a portal to the human proteome for the global research communities and greatly expedite the HUPO international effort.

**Decision support toolkit for climate resilient seaports (AP35)**

This project extends existing applied research activity funded by NCCARF (Enhancing the resilience of seaports to a changing climate), and adds value through the development of an online decision support toolkit which will enable Australian port authorities to make better informed climate risk management decisions. This new ‘applications’ activity will involve three discrete components: the sourcing, refinement, and standardisation of multiple data sets needed for context specific climate change adaptation decision-making, the integration of both primary and transformed data (as produced by new extensible models being developed for the NCCARF project) within a suitable data management framework, and 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 will be an online decision support toolkit which can be readily used by port authorities both in Australia, and potentially internationally. This applications project adds value by building on scientific knowledge, translating science into usable information for the policy community, and through direct application by practitioners.
10.2.6 Applications Project Descriptions - Projects under discussion but not yet contracted as of 30 June 2012

Adaptations and Impacts Downscaling tool (AP04)
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 will build tools that identify suitable downscaled data sets and read the contents of those data sets. A dialog with four key impacts research groups will be established to define their data needs for their research. The tools will be developed to generate these data, as required by the impacts researchers. The data will be exported in formats defined by the impacts researchers to enable their research. All tools, and all generated data sets will be made open source, hosted on an RDSI node and be citeable. We anticipate multiple spin-offs resulting from this project. These include tools usable by the climate community to interrogate climate data sets hosted at an RDSI node. It includes the generation of key data sets required by future downscaling projects. It also includes enabling 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.

Extreme Weather and Health Impacts (AP07)
The project will develop software that enhances existing research data infrastructure with tools that merge population, health and environmental data for analysis and inference in environmental epidemiology. It will build a system for the kind of data sourcing, manipulation and analysis that is required for greater efficiency in many epidemiological research areas. The first demonstration of value will be to publish specific workflows online that integrate datasets and produce an analysis assessing and projecting the health impacts of extreme weather events. The second demonstration of value will be the publication of peer reviewed journal articles that describe the results of our studies. The enhanced infrastructure will eventually provide support for a diversity of epidemiology researchers (e.g., biosurveillance, wildlife health, and emerging infectious diseases). The integrated datasets, workflows and their metadata will be open source, extensible and digitally identified for publishing to Research Data Australia, making the system available and reusable to the wider research community.

Founders and Survivors : Genealogical Connections (AP20)
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 will develop open-source software infrastructure based on Yggdrasil. Each installation of the extended Yggdrasil database application will enable historians to record and analyse the genealogy of an entire particular population group of interest as asserted from birth, death, marriage and other events. Importantly, the evidentiary sources for all asserted events will also be recorded in the database. The project will extend 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. Core research questions initially centre around mortality, health demographics, and life success factors e.g. causes of death, determinants of longevity, family support structures.

Data Transformation and Model Calibration System (AP29)
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) and led by Associate Professor Beverley Henry (Queensland University of Technology) aims to train a number of predictive models using Australian data and test their performance when addressing specific research questions. As a first step, the working group is collating carbon 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 proposes to build a data transformation and model calibration system that will automatically or semi-automatically extract and transform historic data from multiple Australian sites and use them to calibrate and validate the predictive power of multiple carbon and nitrogen dynamics models. The software will capture the process in scientific workflow which will form the blueprint for future similar activities to calibrate and compare ecosystem models using Australian data.

SEQUITOR: A Demonstration Integrated Coastal Knowledge Platform (AP34)
Management across the catchment-coast-ocean continuum remains one of the greatest Natural Resource Management challenges still to be addressed in Australia. This project would develop data integration and visualisation infrastructure for a proof-of-concept Coastal Knowledge platform for South-east Queensland known as SEQUITOR. SEQUITOR will integrate 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. Specific Objectives of the project are: 1). Expose and integrate disparate datasets from a number of organisations in South-east Queensland. 2) Visualise and analyse disparate data in a publically accessible web environment. 3) Assimilate data from observational networks to better constrain catchment and receiving environment models.
## 10.3 Progress against activities

### 10.3.1 Data Capture Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANDS has either entered into contracts (or has substantially agreed on project descriptions) for Data Capture projects at 24 institutions. As noted above, some projects have been, or will be completed before the planning period. Shaded projects are still being defined at time of writing, but are expected to be agreed by the beginning of the planning period.</td>
<td>At June 30, 2012, ANDS had either entered into contracts for all of the Data Capture projects at all of the EOI institutions. This totalled 70 projects, of which 33 are now complete. A breakdown of the progress made in relation to this is provided in section 2.2.2, with projects completed in the current reporting period described in section 2.2.3</td>
</tr>
<tr>
<td>The remaining funds in the Data Capture program will be allocated according to two criteria:</td>
<td>Excess funding was redirected to other ANDS programs, in consultation with the Steering Committee.</td>
</tr>
</tbody>
</table>
### 10.3.2 Research Metadata Store Infrastructure

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

### 10.3.3 Public Sector Data Access Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANDS will commission a fast initial rollout of the SISS offering to an initial group of institutions (that includes several Departments of Primary Industry with whom AuScope is already engaged), who hold well-managed spatial data and are willing to share it.</td>
<td>Engagements with Bureau of Meteorology and GeoScience Australia in progress. Extension of time granted to allow the project to further engagement with the BoM. Relationship with VeRSI established and training and transfer of knowledge has taken place. VeRSI is working with the groundwater community in Victoria – DPI, Vic; DSE, Vic and University of Ballarat.</td>
</tr>
<tr>
<td>ANDS will work with the CSIRO to identify and</td>
<td>Project complete. Murray Darling Basin</td>
</tr>
<tr>
<td>Proposed Activity</td>
<td>Progress</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>expose for discovery an initial set of data collections.</td>
<td>Sustainable Yields data exposed through an institutional portal and RDA. Software deliverables shared through SourceForge and used as the basis for building a CSIRO enterprise data management and publication system.</td>
</tr>
<tr>
<td>ANDS will also fund the development of the National Criminal Justice Data Network to aggregate and provide increased access to Criminal Justice data from a range of state-based data holders in this area, including NSW Bureau of Crime Statistics and Research, NSW Dept of the Attorney-General, the SA Office of Crime Statistics and Research, SA Dept of the Attorney-General, the NT Department of Justice, the Australian Institute of Criminology, and the ABS National Centre for Crime and Justice Statistics.</td>
<td>Data hosted on ASSDA (ADA) and made available via <a href="http://ada.edu.au/crime-and-justice/home">http://ada.edu.au/crime-and-justice/home</a>.</td>
</tr>
<tr>
<td>The selection of further partnerships to make research data discoverable through the ARDC will be based on the public sector data holders’ interest in making data available, their commitment to ongoing data services beyond the life of the project, and the importance of the data to Australian researchers. Consequently selection will occur through discussion with both the data holders and representatives of research disciplines and capabilities.</td>
<td>AODN – project complete. Records from project Underway made available via AODN portal and Research Data Australia. Additional data from AODN portal also harvested. AustLII – Over 400 datasets now available via RDA as well as directly from AustLII. Extension granted to seek further datasets. Powerhouse Museum – Over 1000 collections from 18 museums from around Australia are in the final stages of quality assurance before exposure via RDA. NAA – Proposal to identify collections and feed into ARDC has received no response and other opportunities will be sought to expose this data. GeoScience Australia/OSDM – Collaborative work to map key data</td>
</tr>
</tbody>
</table>
### Proposed Activity

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>catalogues for exposure of an initial 7000 records into ARDC is complete. This was an ‘assisted engagement’ where ANDS staff go on site and work with GA staff. It is awaiting some IT development within GA to allow the feed to be activated.</td>
<td></td>
</tr>
<tr>
<td>Aus-e-Lit – Annotation service and compound authoring tool for the Australian literature community. This project is in the final stages.</td>
<td></td>
</tr>
<tr>
<td>DIAS-B – Provision of annotation service for Atlas of Living Australia. This project is also in the final stages.</td>
<td></td>
</tr>
<tr>
<td>SISS – development of spatial data infrastructure for AuScope to build new spatial data services out of existing ones. Commencing with GA and BoM data using NCI.</td>
<td></td>
</tr>
</tbody>
</table>

### 10.3.4 Australian Research Data Commons Core Infrastructure

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Data Collections Registration Infrastructure:</strong></td>
<td>Multi-format support re-scheduled to 2013</td>
</tr>
<tr>
<td>Multi-format support: Pre-harvest cross-walk</td>
<td>Admin Interfaces updated</td>
</tr>
<tr>
<td>Admin interfaces</td>
<td>Architecture for third party deployment complete</td>
</tr>
<tr>
<td>Architecture for third party deployment</td>
<td>Component architecture commenced</td>
</tr>
<tr>
<td>“Open source” management</td>
<td></td>
</tr>
</tbody>
</table>

<p>| <strong>National Data Collection Discovery Infrastructure:</strong> | See Also DataCite completed |
| Provide further ‘See Also’ linking | Browse by ANZSRC FOR and filtering by license completed |
| Allow further structured searching | Scaling commenced |
| Scaling | Architecture for third party deployment |</p>
<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture for third party deployment</td>
<td>complete</td>
</tr>
<tr>
<td>“Open source” management</td>
<td>Component architecture commenced</td>
</tr>
<tr>
<td><strong>National Data Collection Description Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>Update information and presentation design</td>
<td>Belvedere project completed</td>
</tr>
<tr>
<td>Scaling</td>
<td>Scaling commenced (search index decoupled)</td>
</tr>
<tr>
<td>Architecture for third party deployment</td>
<td>Architecture for third party deployment</td>
</tr>
<tr>
<td>“Open source” management</td>
<td>Component architecture commenced</td>
</tr>
<tr>
<td><strong>Dataset Identifier Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>Integrate the ANDS DOI service with existing ANDS services</td>
<td>Production DOI service released</td>
</tr>
<tr>
<td>Continue to contribute to the design and implementation of global data identifier and citation system</td>
<td>Contributions made to DataCite working groups and meetings</td>
</tr>
<tr>
<td></td>
<td>First global use of DataCite search API</td>
</tr>
<tr>
<td><strong>Place Names Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>ANDS is in partnership with responsible government agencies (GA/OSDM) to establish a robust national infrastructure that will allow place names to be validated by both individuals and software systems against an Australian Gazetteer (a directory that lists names of geographical place and features and includes spatial co-ordinates)</td>
<td>Gazetteer service launched</td>
</tr>
<tr>
<td>Phase 2 (marine, boundary boxes, expanded dataset)</td>
<td>Phase 2 (boundary boxes, expanded dataset and transactional web services)</td>
</tr>
<tr>
<td>Phase three planning (historical names, vernacular names, crowd sourcing...)</td>
<td>Complete</td>
</tr>
<tr>
<td>Integrate Gazetteer services with existing ANDS services</td>
<td>Phase three planning commenced with CGNA (historical names, vernacular names, crowd sourcing...)</td>
</tr>
<tr>
<td></td>
<td>Integrate Gazetteer services with existing ANDS services re-scheduled for Q4 2012</td>
</tr>
<tr>
<td>Proposed Activity</td>
<td>Progress</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Researcher Identification Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>Establish Infrastructure with National Library of Australia</td>
<td>Party infrastructure launched</td>
</tr>
<tr>
<td>Project Benefit Realisation stage</td>
<td>Project benefit realisation commenced</td>
</tr>
<tr>
<td>Integrate Researcher Identification Infrastructure with existing ANDS services</td>
<td>Integration workflows complete; implementation scheduled</td>
</tr>
<tr>
<td><strong>Research Activity Information Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>In collaboration with the Australian Research Council and the National Health and Medical Research Council create software components and services that will improve the accessibility and quality of information about research activities undertaken within Australia. When completed, the project will enable the public to discover and track the research undertaken through ARC and NH&amp;MRC funding and to follow its connections to research outputs, both nationally and internationally.</td>
<td>Slippage; funding body timetables incompatible; project paused</td>
</tr>
<tr>
<td><strong>Standard Vocabularies Infrastructure:</strong></td>
<td></td>
</tr>
<tr>
<td>Develop and implement software to enable the creation, management, and publication of human and machine readable ‘terminologies’ (also known as controlled vocabularies) for use by the Australian innovations sector.</td>
<td>Project de-prioritised</td>
</tr>
<tr>
<td>Project de-prioritised</td>
<td>First phase complete: internal vocabulary service drives browse by ANZSRC feature in Research Data Australia</td>
</tr>
<tr>
<td><strong>ARDC Infrastructure Establishment:</strong></td>
<td></td>
</tr>
<tr>
<td>Provide technical assistance and design support to those organisations building the distributed infrastructure of the ARDC</td>
<td>Specialist technical support provided through ANDS service desk, specialist DOI infrastructure implementation support, specialist Party Infrastructure implementation support, and participation in webinars, events and developer “dojos”;</td>
</tr>
<tr>
<td>Provide specialist technical backup to members of the Australian research and higher education sector in the uptake of ANDS infrastructure</td>
<td>Maintenance of ANDS service desk and associated systems and processes</td>
</tr>
<tr>
<td>Continue the establishment of ANDS infrastructure to support those services which have been</td>
<td>Participation in the ANDS international</td>
</tr>
</tbody>
</table>
### Proposed Activity

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>implemented to date</td>
<td><strong>Progress</strong></td>
</tr>
<tr>
<td>Continue to work to enable integration of ARDC Core Infrastructure with international infrastructure networks (such as NSF and JISC), including the registration of Australian data catalogues in Research Data Australia</td>
<td>engagement program</td>
</tr>
</tbody>
</table>

---

#### 10.3.5 **Australian Research Data Commons Applications Infrastructure**

<table>
<thead>
<tr>
<th>Proposed Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the area of data integration/fusion/merging, the ARDC project will fund AuScope through CSIRO to build new spatial data services out of existing services. The project will provide the Atlas of Living Australia with the ability to use annotation to improve data quality; and to integrate a variety of literary data sources for improved discovery. Lastly, the project will integrate a large number of sources of data associated with the coast from many disciplines to enable exploration of the data through common metadata and discovery tools.</td>
<td>AuScope: ANDS is funding a NeAT project to develop the Spatial Information Services Stack (SISS) Atlas of Living Australia: ANDS is funding the DIAS-B NeAT project. Coast: ANDS took part in a series of discussions and workshops to identify a possible activity related to coastal data. As yet, this has not led to anything.</td>
</tr>
<tr>
<td>In the area of data visualisation, the ARDC project will commission a tool that enables the visualisation of the connections between projects, researchers, and their research data to show the relationships between them.</td>
<td>This activity has not yet taken place.</td>
</tr>
<tr>
<td>In the area of data analysis, the ARDC project will initiate an activity with the University of Queensland as candidate service developer and a service provider (to be identified) to mark up existing data with a Who/What/When/Where data combination service and demonstrate new analysis opportunities on top of this marked-up data. All of these activities have target owners and operators of the resulting services.</td>
<td>A report on the options for providing such a service is about to be delivered.</td>
</tr>
</tbody>
</table>
10.4 Risk Register

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

10.4.1 Political and Governance

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

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

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

Risk 2 – That the ANDS Project is not managed effectively

Risk Factors:
- Lack of effective mechanisms for planning, leadership and management.
- The structure of ANDS has a negative impact on coordinated delivery of required activities.
- Collaboration between the Project and across locations is not effective.
• ELF funding guidelines do not allow for sufficient Project staff to administer funded programs of work.
• State based staff have mixed allegiances.
• Projects starting too late.

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 all projects commence by June 2012.
• Ensure all late starting projects are closely managed.

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

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

Risk Mitigations:
• Approval has been obtained for stream-lined approaches at Monash University to enable ANDS to work more effectively.
• Fund additional staff or specific work at Monash University to enable ANDS to work more effectively.

10.4.2 Relationships

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

Risk Factors:
• Stakeholders are not prepared to undertake the changes within their own organisations that are necessary for the realisation of the Australian Research Data Commons.
• Stakeholders do not see their interests in data management and those of the Project as being aligned.
- Stakeholders might feel that the wrong decisions have been made.
- ARDC Applications program - Stakeholders might feel that the wrong decisions have been made in determining which projects to fund.
- Substantial new funding for other eResearch activities and reduction in new funding for ANDS projects.

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

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

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

**Risk Mitigations:**
- 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.
- Formal procurement processes have been implemented to ensure that the requirements are understood and that potential suppliers meet the set criteria.
Risk 6: That ANDS is not perceived as a long-term partner and hence our services are not taken up

Risk Factors:
- The impending end of ANDS NCRIS and EIF 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.
- Strong contribution to DIISRTE Roadmap process will be a mitigating factor.

Risk 7: That there is confusion about the role of ANDS versus other related service providers in the eResearch sector which impedes effective service delivery

Risk Factors:
- ANDS and PfC partners’ offerings are confused by possible users.
- Relationship between ANDS and state-based eResearch providers (such as Intersect) is not clear to users.

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 e-Research 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.

10.4.3 Impact

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

Risk Factors:
- The storage needs of researchers are not met, so they 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:**

- 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.
- Recommend that funding be linked to the provision of data via the ARDC as it becomes available.
- Ensure that ethics agreements balance confidentiality with openness.
- Provide targeted assistance to data federations to assist with integration into the ARDC.

**Risk 9 – That reusers 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 reuse 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:**

- Seek international engagements and partnerships to take up standards and technologies favoured by ANDS and share development load.
- Ensure enough people are trained on the standards and technologies that ANDS is adopting to support wide adoption.
- Make implementation decisions such that ANDS is not dependent on particular standards and technologies, but on general approaches that can be transferred across technologies.
- Encourage the use of ANDS-developed technologies by other data aggregators such as Terrestrial Ecosystem, Research Network (TERN).
10.4.4 Resourcing

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

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

Risk Mitigations:
- Provide good information to staff of likely outcomes beyond the current funding.
- Build a vision for the function of ANDS for the longer term and communicate this to staff.