AUSTRALIAN NATIONAL DATA SERVICE (ANDS)
National Collaborative Research Infrastructure Strategy (NCRIS)

Progress Report 3, 30 September 2011
# TABLE OF CONTENTS

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

2# ACTIVITIES UNDERTAKEN..................................................................................................... 10#
   2.1# Research Infrastructure ................................................................................................. 10#
   2.2# Seeding the Commons ................................................................................................... 10#
   2.3# Frameworks and Capabilities ......................................................................................... 18#
   2.4# Project Office ................................................................................................................ 26#
   2.5# Promotion ..................................................................................................................... 27#
   2.6# Risk Management .......................................................................................................... 29#

3# PROGRESS AGAINST MILESTONES................................................................................... 31#
   3.1# Seeding the Commons ................................................................................................... 31#
   3.2# Frameworks and Capabilities ......................................................................................... 32#

4# DEVIATIONS FROM THE PROJECT PLAN........................................................................... 34#

5# CO-INVESTMENT .................................................................................................................... 36#
   5.1# Access and Pricing ......................................................................................................... 36#
   5.2# Project Co-Investment .................................................................................................... 36#

6# PERFORMANCE INDICATORS.............................................................................................. 37#
   6.1# KPI Report .................................................................................................................... 37#
   6.2# Progress over the Life of ANDS Projects .................................................................... 40#

30 September 2011
6.3# Overall Progress.................................................................41#

7# APPENDICES........................................................................43#

7.1# Confidential Information ......................................................43#

7.2# Project Description Detail ......................................................43#

7.3# Progress against activities....................................................53#

7.4# Risk Register .......................................................................59#
1 Project Status

1.1 Background

Research data is steadily becoming more voluminous, more complex, and more important. The nature of research is changing. It has become more investigative as it is possible to assemble significant data collections that enable much broader problems to be addressed. Thus it is very important 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 thus 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 having more researchers re-using 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 has 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 2009-10 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 2010-11 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; and
- Building Capabilities - improving Australia’s capability to manage its research data.

As a result of the ARDC project, the NCRIS ANDS project consolidated to 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; and
- Seeding the Commons – improving local data capture and populating the data commons.

The associated ARDC project has five programs of activity:

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

30 September 2011

ands.org.au
3. **Metadata Stores** – an institutionally based program that enables metadata to be stored coherently across an institution that supports data management, publishing, sharing and re-use

4. **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.)

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

These programs are complemented and enveloped by the two NCRIS programs, *Seeding the Commons* and *Frameworks and Capabilities* as shown in Figure 1.

![Figure 1: Relationship between Programs](image)

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

As a result of the first few years of ANDS activities many Australian researchers in a wide range of disciplines across Australia’s research institutions are able to:

- capture data, with rich metadata,
- automatically store it,
- simultaneously publish through a discipline portal and Research Data Australia,
- integrate that data with other data discovered through a portal, and
- publish both the results and the data of their investigations.

Consequently, the Australian Research Data Commons (ARDC) has been established.

The ARDC is a combination of the set of shareable Australian research collections, the descriptions of those collections including the information required to support their re-use, the relationships between the various elements involved (the data, the researchers who produced it, the instruments that collected it and the institutions where they work), and the infrastructure needed to enable, populate and support the commons. ANDS does not hold the actual data, but points to the location where the data can be accessed. The ARDC can be envisaged below, where ANDS is contributing to the green pipes and boxes:

![Diagram of Australian Research Data Commons](ands.org.au)

**Figure 2: Australian Research Data 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. Substantial progress has been made. In Figure 3: Australian Research Data Commons Progress, we will describe the composite effect of the activity of ANDS to date, including this year.
Secondly ANDS has provided a meeting place that enables research data management to be progressed for the whole of Australia. Beyond establishing this interchange, ANDS also enables a coherent Australian response to the rise of research data, and has enabled all relevant participants to engage. ANDS is engaged with all major research institutions, and importantly they are engaged with each other. ANDS is engaged with major research data collection agencies, such as GeoScience Australia and the Australian Bureau of Statistics. ANDS is engaged with the data intensive infrastructure partners such as the Integrated Marine Observing System, the Australian Telescope, AuScope, and the EBI mirror. ANDS is engaged with funding agencies, principally the ARC and the NHMRC, and ANDS is engaged internationally with national counterparts such as the UK’s JISC, the NSF in the USA and the Netherland’s SURF, as well as international initiatives that are developing between Europe and the USA.

The major activities for ANDS as a whole in the 2010-11 calendar year were:

- Enabling research data collections to be described and harvested by ANDS at 26 research institutions
- Enabling research data collections to be described and harvested by ANDS at 27 research data providers including other research infrastructure providers, public sector providers, cultural institutions and research consortia
- Making 26,746 research data collection pages from 21 collections providers discoverable through Research Data Australia, Google, and other search engines
- Establishing infrastructure to identify, register and publish collections descriptions through Research Data Australia, the ANDS portal into the Australian Research Data Commons
- Helping to establish a coherent 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
- Concluding the NeAT program of discipline enhancing tools for improved collaboration and exploitation of research data

There have been many highlights over the past year. They can be described as the establishment of national services, the establishment of coherent institutional research data infrastructure, using all of this infrastructure to enable the population of the commons, and enhance the ability of the research system to exploit this improved research data environment.

Establishing national services:

- Enhanced ANDS data collections registration service,
- Enhanced ANDS collection description publication service,
- Enhanced Research Data Australia – a data collections discovery service,  
- A dataset identification service established, and
- Party Identifier Infrastructure service run by the NLA enabling richer collections descriptions in the commons

Helping to develop coherent institutional research data infrastructure:
- Tools have been deployed to automatically capture rich metadata along with the data for a wide range of instruments
- 9 institutions operate enterprise-supported metadata stores with 15 more under development
- 4 institutions have operational research data management plans and with 15 more under development
- 30 institutions have installed services that enable collections information to be harvested

Populating the Commons:
- 26,746 collections were described and available through Research Data Australia by June 30th, 2011
- Over thirty research data providing institutions were engaged through direct projects and our partnerships with AuScope and the Museum Metadata Exchange projects,
- 22 institutions will provide collections descriptions feeds to ANDS, both research institutions and public sector data holders, and
- 2 discipline oriented portals were cross connected to Research Data Australia.

Enhanced ability to exploit the research data environment:
- The majority of the tools developed in NeAT projects concluded an delivered benefit to researchers in the relevant domains
- Improved licensing regime supported through AusGoal enabling simpler use and data integration
- International initiatives to ensure compatible approaches are adopted to research data internationally

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 was to enable research institutions to develop and exploit their research data ambitions whilst meeting the goals of Seeding the Commons and Data Capture programs.
- The pace of engagement was slower than planned, but appears to be delivering stronger outcomes as a result of deeper engagement.
- ANDS has tested its national approach to collections that augments disciplinary approaches internationally. This approach has found to be internationally valuable but able to interact with other jurisdictions that take other approaches.
- ANDS decided that in order to have researchers change their approach to research data and its management, the optimal path was through the research institutions. This approach has continued to be strongly supported by the institutions, and not resisted by researchers.
- The ANDS partner agreement and steering committee was shown to be resilient to a significantly changing environment.

To now summarise the progress of the establishment of the ARDC, we show the many components of the ARDC that have been completed by 30 June 2011 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.

![Diagram of Australian Research Data Commons Progress](ands.org.au)
2 Activities Undertaken

2.1 Research Infrastructure

ANDS has continued to make progress towards its goals of providing greater support to enable researchers to work in the new world of data-intensive research. More detailed reports on progress in this area are contained in UUs 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.

Much of this infrastructure is being established as part of the EIF ARDC 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. However, the infrastructure that enables the effective use of the ARDC is the key focus of the ANDS project – the institutional ability through policy, procedures, people, systems to engage with research data, whether those institutions are the research institutions or part of the wider research data community – data providers, data beneficiaries, policy organisations, public sector organisations and funding bodies.

The infrastructure has been constructed through two programs that complement the ARDC programs:

- **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; and
- **Seeding the Commons** – improving local data capture and populating the data commons.

The next section describes the specific research infrastructure created in the 2010-11 financial year.

2.2 Seeding the Commons

2.2.1 Overview of program

The aims of this program are to improve the fabric for data management in a way that will increase the amount of content in the data commons; and to improve the state of data capture and management across the research sector, with a focus on the tertiary education sector, CSIRO and the NCRIS Capabilities.

ANDS has a significant existing set of commitments for this program. Its major activities are:

- 34 research institutions developing local capability for managing research data and making collections visible in the ARDC
- Regional and NCRIS capability local support through funded positions
- National advice and direct support

These activities are focused on growing the commons, supporting partners, and providing national advice. Activities to grow the data commons will provide:
• advice on metadata standards and requirements to ensure that metadata is prepared in a manner consistent with ANDS’ needs
• advice on ANDS service requirements to ensure that metadata is available in ARDC
• funding of staff at partner projects and institutions to provide support for ANDS’ goals
• identification of as much content as possible and making it discoverable

2.2.2 Outline of projects
An analysis of research intensity for the major Australian research-producing institutions was undertaken in late 2009 based on the most recent publicly available data, and $4.55M of Seeding the Commons funds were allocated in bands of $250K, $125K, or $75K, as detailed below. Institutions were sent an invitation to take part in an Expression of Interest process in late 2009.

<table>
<thead>
<tr>
<th>Institution</th>
<th>$ ,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSIRO</td>
<td>250</td>
</tr>
<tr>
<td>The University of Sydney</td>
<td>250</td>
</tr>
<tr>
<td>The University of Melbourne</td>
<td>250</td>
</tr>
<tr>
<td>The University of Queensland</td>
<td>250</td>
</tr>
<tr>
<td>University of New South Wales</td>
<td>250</td>
</tr>
<tr>
<td>Monash University</td>
<td>250</td>
</tr>
<tr>
<td>The Australian National University</td>
<td>250</td>
</tr>
<tr>
<td>The University of Adelaide</td>
<td>125</td>
</tr>
<tr>
<td>The University of Western Australia</td>
<td>125</td>
</tr>
<tr>
<td>University of Wollongong</td>
<td>125</td>
</tr>
<tr>
<td>Queensland University of Technology</td>
<td>125</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>125</td>
</tr>
<tr>
<td>Griffith University</td>
<td>125</td>
</tr>
<tr>
<td>Curtin University of Technology</td>
<td>125</td>
</tr>
<tr>
<td>Royal Melbourne Institute of Technology</td>
<td>125</td>
</tr>
<tr>
<td>University of South Australia</td>
<td>125</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>125</td>
</tr>
<tr>
<td>University of Technology, Sydney</td>
<td>125</td>
</tr>
<tr>
<td>La Trobe University</td>
<td>125</td>
</tr>
<tr>
<td>Deakin University</td>
<td>125</td>
</tr>
<tr>
<td>University of Western Sydney</td>
<td>125</td>
</tr>
<tr>
<td>University of Tasmania</td>
<td>125</td>
</tr>
<tr>
<td>University Name</td>
<td>Funding</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>The Flinders University of South Australia</td>
<td>125</td>
</tr>
<tr>
<td>James Cook University</td>
<td>125</td>
</tr>
<tr>
<td>Charles Sturt University</td>
<td>75</td>
</tr>
<tr>
<td>Victoria University</td>
<td>75</td>
</tr>
<tr>
<td>University of New England</td>
<td>75</td>
</tr>
<tr>
<td>Swinburne University of Technology</td>
<td>75</td>
</tr>
<tr>
<td>Edith Cowan University</td>
<td>75</td>
</tr>
<tr>
<td>Murdoch University</td>
<td>75</td>
</tr>
<tr>
<td>University of Canberra</td>
<td>75</td>
</tr>
<tr>
<td>University of Southern Queensland</td>
<td>75</td>
</tr>
<tr>
<td>Central Queensland University</td>
<td>75</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$4,550</strong></td>
</tr>
</tbody>
</table>

Charles Sturt University did not respond to the EoI. The funds allocated to them will be used on other areas of the program, notably ANSTO. Those universities that did not receive funding as part of this program will be offered support and advice by program staff, which has already occurred with University of the Sunshine Coast and University of Ballarat. The University of Ballarat is also engaged in a project to make borehole data available through another project.

At June 30, 2011, ANDS had either entered into contracts (or had agreed on project descriptions) for Seeding the Commons projects at all of the EOI institutions with the exception of the University of Western Sydney and the University of Southern Queensland. A breakdown of the progress made in relation to this is provided in section 2.2.3. Detailed descriptions of the contracted and agreed projects can be found in Section 10.2.

Other activities to work more closely with partners have included:

- review and assessment of partner projects to ensure they are completed on time and as specified
- advice to these projects and other ANDS programs
- advice and assistance on data management and related policy and procedures
- advice and assistance in the use and deployment of ANDS produced or funded services, applications and material
- identification of and partnering with exemplar institutions to maximise data management

ANDS has sought to create a community of data managers through:

- founding and supporting several state based or related groups of data managers, with effective communication channels between them.
- training provided to Community members as required (in conjunction with Frameworks and Capabilities Program)
- establishing an ANDS Knowledge Bank, for use by the community
- capturing information about successes for dissemination within the community and beyond developing relationships with equivalent activities overseas to share approaches to data management systems that can inform ANDS

### 2.2.3 Activity/Deliverables for 2010-11

Seeding the Commons’ primary focus has been in engagement with ANDS’ partners and with research producing institutions. This has involved active engagement and interaction with the 34 institutions identified during the EOI process, to assist them in not only defining their projects, but also in undertaking them. Seeding the Commons staff have been advising partner institutions in areas such as metadata, automated feeds, use of ANDS tools (such as the Research Data Australia “Sandbox”) and engagement strategies for working with researchers. This has also entailed working closely with the Capabilities program within ANDS to provide documentation and guides.

Unlike the Data Capture program, the Seeding the Commons projects tend to be similar in nature. Common elements of these projects include:

- Identification of existing data sets within an institution, with a view to enabling the expose and description of them were possible
- Examining existing processes, policies and workflows, in data management with the intention of using what is learnt to improve data management capabilities within the institution
- Creating descriptions of datasets to increase understanding of how these descriptions can best be created
- Building tools to enable and simplify metadata capture related to datasets

The following table demonstrates the current status of the projects in this area. Descriptions of the projects underway or completed are listed in [section 10.2](#).
<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>ANSTO</td>
<td>Scientific Information Architecture</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Australian National University</td>
<td>The ANU Seeding the Commons Project</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Queensland University</td>
<td>CEM Core Data Curation project</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CSIRO</td>
<td>Seeding the Commons: Enabling CSIRO's Biological Collections for the ARDC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Curtin University of Technology</td>
<td>Researcher Centric Model for the management of Research Data</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deakin University</td>
<td>Description and discovery of research data collections available at Deakin University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Edith Cowan University</td>
<td>Data Management Plan and Policy</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Flinders University</td>
<td>Reformatting the AusStage dataset to support access and re-use by researchers AND Reforming the Movies: the Motion Picture Producers and Distributors of America, Inc. database</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Griffith University</td>
<td>Identifying and describing Griffith University's research datasets and making metadata available to the Griffith research community and the Australian Research Data Service</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Griffith University/Macquarie University</td>
<td>National Linguistics Corpus</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>James Cook University</td>
<td>Systematic JCU Tropical data</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

30 September 2011

ands.org.au
<table>
<thead>
<tr>
<th>University</th>
<th>Project Description</th>
<th>X</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>La Trobe University</td>
<td>Archaeological Database Development: The People and Place Project</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Macquarie University Seeding the Commons</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Monash University</td>
<td>Monash University Seeding the Commons Project</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Murdoch University</td>
<td>Integrating precision agriculture</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Queensland University of Technology</td>
<td>Seeding the Commons Funding</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RMIT University</td>
<td>Screen Media Research Archive</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Swinburne University of Technology</td>
<td>Swinburne’s “Watering the garden for the seeds to grow” project</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>The University of Melbourne</td>
<td>Seeding the Commons Funding</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>The University of Queensland</td>
<td>Seeding the Commons Funding</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Adelaide</td>
<td>Research Data Storage and Management</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Canberra</td>
<td>Cross-communication and enhanced accessibility for research data management systems</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of New England</td>
<td>UNE’s N.C.W. Beadle Herbarium Database</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of New South Wales</td>
<td>Research Data management Services</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>Newcastle Research Data Online (Seeding the Commons)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of South</td>
<td>Taking Australian</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

30 September 2011

ands.org.au
<table>
<thead>
<tr>
<th>University</th>
<th>Action</th>
<th>Australia</th>
<th>University of Southern Queensland</th>
<th>University of Sydney</th>
<th>University of Tasmania</th>
<th>University of Technology Sydney</th>
<th>University of Western Australia</th>
<th>University of Western Sydney</th>
<th>University of Wollongong</th>
<th>Victoria University</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Architectural and Built Environment Records into the Commons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sustainable policy and procedure for capturing research data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seeding the Commons at the University of Sydney</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Publication of collections into the ARDC by UTAS</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Community Tools and Processes for Effective Data Management Planning</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Seeding the Commons through research data management at the University of Western Australia</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>UWS Seeding the Commons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identifying and locating UOW data sets to seed the Australian Research Data Commons and the development of a supporting research data management policy</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Research data framework</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Most of these projects were still underway in the reporting period, however many have reached early milestones, including the delivery of records into Research Data Australia, which has grown substantially during the reporting period, increasing the amount of data discoverable and accessible through the data commons.

Members of the Seeding the Commons team have also been working to provide engagement support for the Public Sector and Data Capture programs which have smaller staffing compliments, and no presence outside of the ACT and Victoria.

The program was able to do this due to the provision of outreach staffing. ANDS has entered into contractual agreements to employ staff at eResearch South Australia, iVEC (WA), CAIRSS (Qld) and Intersect (NSW).
These staff enable ANDS to have an active presence in cities without an ANDS office. ANDS has also actively partnered with these organizations in the running and administration of projects.

Program staff have also engaged with institutions that were not part of the EOI program (such as University of the Sunshine Coast and University of Ballarat), existing relevant support organizations (CAUL Institutional Repositories Support Service or CAIRSS) and with disciplinary based groups to provide other support. An engagement with linguists based at Griffith and Macquarie universities has involved discussion on a project for a National Linguistics Corpus which is underway.

Other activities have focused on building the community of data managers needed to support ANDS’ goals into the future. ANDS ran a series of Community Events in the first half of 2011, as part of our ongoing community building efforts. These were held in Brisbane, Perth, Adelaide, Melbourne and Sydney, with a total of 193 partner staff in attendance. These partners presented on 92 different ANDS funded projects.

Feedback from partners indicated that they found hearing about other projects, making connections with others, and learning from other partners and ANDS staff most valuable. Having various experts in the room also meant that partners were able to seek clarification on issues that were specifically relevant to their project. During each event partners broke into discussion groups to either problem solve or share issues. These discussions have been captured by ANDS to help prioritise future support.

To further build a community of data management experts ANDS plans to host more events over the next two years, with a focus on connecting those with specific roles and needs in common, for instance for data librarians, for technical staff, for those interested in ontologies, etc. At the end of each event partners were encouraged to join the ands-partners Google group and the Community Bulletin Board, to help further share information.

### 2.2.4 Program Highlights, Issues and Breakthroughs

A particular highlight to date has been the growth in awareness of the importance of institutional support for data management as a result of these projects. In many cases the exemplar projects in particular began in isolation (within a division or research group) but have created links across the institution and have resulted in new alliances and a desire to internally funded needed infrastructure such as storage and data management professionals.

The establishment of the National Linguistics Corpus project, which is a collaboration between Griffith and Macquarie Universities, and which has the support of the Australian Academy of Humanities, Australian Linguistics Society, the ARC Network in Human Communication Science and a consortium of Australian universities, offers a promising insight into the way that future National Collections of data might be organised, as well as integrating a number of strategic data sources.

The very well attended series of community events in the first half of 2011 also demonstrated the interest in ANDS partners working together, and established the underpinnings of an Australian community of data managers.

30 September 2011

ands.org.au
2.2.5 Program Learnings

It had been expected that the metadata requirements for adding records to Research Data Australia were relatively straightforward and would not pose many delays, especially if metadata experts such as librarians were involved. However it has become clear that despite ANDS having offered some training and provided extensive documentation, many of our partners were struggling with the concept. ANDS also had limited in-house expertise to advise on and assess records. This was further complicated by changes to our standards as we learnt more, and by the introduction of new services, such as the People Identifier service.

Learning: ANDS needed to provide more advisors and assessors, which was done in the first half of 2011, and offer more training, which is being done as a combination of improved documentation and more sessions. Communication around changes needed to be improved as well, which has occurred.

Overall, everything took longer than expected: discussions on projects took time to converge on a set of activities that met the institution’s desires and ANDS’ requirements, negotiations between institutional legal staff often took time, even with standardised contracts, and the finance processes of the lead agent introduced delays of between one and two months before funds flowed to the institutions. Once funds were in place recruitment was often required at partners, which was often difficult, and in some cases the recruitment was too focused on software needs, rather than the requirements of describing data collections. Assessment of deliverables by ANDS was too slow for many partners, and impacted on payments.

Learning: Despite many improvements to internal processes many projects could only be agreed through extensive face to face contact to ensure that the partner gave it their full attention. More resources were added to ANDS for metadata assessment, and the assessment process in general was reviewed and improved. The addition of a dedicated contracts manager to the team greatly improved the contracting and payment processes.

2.3 Frameworks and Capabilities

2.3.1 Overview of program

The Frameworks and Capabilities program is an attempt to address two of the systemic obstacles to the emergence of the ARDC: policy irregularity/absence and human capability constraints.

The common approach to addressing both of these generic issues is to partner with collaborators around specific solutions. The Frameworks and Capabilities program produces materials that address some of the fundamental shared issues in data intensive research. The key collaborators for the Capabilities activities within the program are eResearch support groups and the key collaborators for the Frameworks activities are research leaders and funding agencies. A central concern of the program as a whole is the desire for research organisations and research groups to have effective policies around the full lifecycle of data management.

The Frameworks activities within the program aim to support new approaches to data-intensive research by strengthening the overall policy context for, and facilitating the emergence of, the ARDC. The Capabilities activities aim to improve the level of capability for data intensive research (and associated technologies)
across Australia by working with research groups and partnering with willing institutions to improve core data competencies.

The Frameworks activities are focused primarily on the research community and the institutions in which they work (as well as the collaborators described below), which include universities, museums, libraries, galleries and government agencies. The activities will work towards harmonising and streamlining the overall policy framework within which a data commons can operate. The result will be a shared vision of the opportunities, benefits and responsibilities of a data commons. It is important to acknowledge that the Frameworks activities work through facilitating the goal of an effective research data commons rather than prescribing the specific policies required to achieve that end.

The Frameworks activities seek to build bridges between researchers, research institutions, research funders, and data creators and curators. In addition, the program will engage with current and emerging initiatives such as the Government 2.0 Taskforce report (in collaboration with the Public Sector Data program) and the National Committee for Data for Science. ANDS primary collaborators include:

- Institutional data holders (CSIRO, NCRIS Capabilities, National Library of Australia, Departments of Primary Industry, GeoScience Australia, Australian Bureau of Statistics, etc)
- National initiatives such as the National Committee for Data for Science
- Cross-governmental groups such as Australian Government Information Management Office (AGIMO) and the Office of Spatial Policy (OSP)
- Research funding departments such as DIISR and the Department of Employment, Education and Workplace Resources (DEEWR)
- Research funding schemes such as the Australian Research Commission (ARC), National Health and Medical Research Council (NHMRC), Research Infrastructure Block Grants (RIBG)
- Discipline leaders within institutions
- Research office staff at institutions

As cohesive networks of research data are increasingly regarded as an important and enduring part of the collaborative research infrastructure, the Capabilities activities focus in particular on building the capability of researchers and support staff to contribute to and better exploit national data infrastructure. The various activities work with the sector to identify and document the fundamentals of working with research data and the specifics of discipline-based data-intensive research. They also work with research communities and local e-Research support services to improve particular data-related competencies, as well as enhancing and adding national focus to institutionally based support, materials development, and training initiatives. ANDS supplements the activities of this program in the community by establishing a reference group drawn from research communities, data networks, e-Research support services, and data stewards. This group assists ANDS to identify the key research data competencies required in the community and identify various means to improve them.

The Capabilities program provides materials and resources for those using ANDS services and in particular supports the needs of ANDS funded project activity.

The Capabilities activities lead to services such as consultancy, informal knowledge transfer, workshops, documentation, and training materials. Staff from this program work within an integrated engagement activity with staff from all other ANDS programs. ANDS will identify and engage the community of
researchers and e-Research support services. These groups themselves are engaged in capability building within their own institutions as well as with their own staff. This engagement is an opportunity to build a community of those who have a stake in building the capability of Australian researchers to contribute to, and get benefit from, national data infrastructure. ANDS expects to inform its own activity through this engagement as well as facilitate some cohesion within the sector.

2.3.2 Outline of projects

The Capabilities area of this program is a well-established program of ongoing activity. A number of projects have been undertaken in the three areas of activity:

Community and Capability Building:
- Establishing ANDS Partners Community, through Google Groups, Community Bulletin Board, and ANDS Community Days
- Conducting Public Events, principally ANDS Roadshows, and an eResearch Australasia 2010 workshop
- Providing targeted training, though Research Data Management “Bootcamps”, specific training as at the University of Newcastle, and conducting a “Gumboots for Data Deluge” workshop

Guides and Support Materials:
- ANDS Content Providers Guide (major project)
- Guides and Topic pages

Ongoing projects have continued and new initiatives have been established in the three activity areas of Frameworks:

Data Commons Policy:
- Submissions to ARC, NHMRC, and DIISR on data-friendly policy settings
- Gov 2.0 policy liaison

Licensing Frameworks:
- AusGOAL for the research sector

Analytical Basis:
- Value of Data project

The strategic re-use of publicly funded data is at the heart of the Australian Research Data Commons and this requires publicly funded data to be made freely available (either at no cost or at marginal cost of provision). Research and public sector organisations need ways of valuing the data itself, valuing their own data provision, valuing the re-use made by others, and communicating that value to the society at large.

The Frameworks program is partnering with publicly funded research organisations and data collection agencies to perform a study to:
- identify the costs involved with data provision
- identify the value of participating in the data commons
By using well-understood techniques, including treasury modelling, the Frameworks program is developing a persuasive set of arguments, based on real case studies, which can be used to encourage the public and innovation sector to participate more fully in the evolving Gov 2.0 and 'Innovation 2.0' agendas.

### 2.3.3 Activity/Deliverables for 2010-11

**Community Building**

ANDS Partners - a Google Group was established in 2010 to facilitate interaction between those engaged in ANDS-funded projects in Australian universities and research institutions. ANDS Partners continues to see lively interaction on a variety of topics. ANDS Partners is also open to others who have an interest in the detail of ANDS activities and services. Discussion regarding the definition of research data resulted in a guide on the subject “What is Research Data?” [http://ands.org.au/guides/what-is-research-data.html](http://ands.org.au/guides/what-is-research-data.html)

Community Bulletin Board – the ANDS Partners Google Group operates in association with the ANDS Community Bulletin Board, which can be seen at [http://community.ands.org.au](http://community.ands.org.au). The Bulletin Board has two parts: a section that is open to the world, and a section where ANDS partners can log in to access shared documents. Capabilities have been actively encouraging more use of the Community Bulletin Board in 2011. This resulted, in one example, of the number of ‘reads’ on the RIF-CS Changes discussion thread increasing from 296 to 1290 in one month (July 2011).

**Capability Building**

*Public Events*

**Roadshows:**

The 9th and final Roadshow was held at James Cook University in Townsville in September 2010. Eight Roadshows were previously held between July 2009 to June 2010, in Adelaide, Brisbane, Canberra, Melbourne, Newcastle, Perth, Hobart and Sydney. A total of over 300 people attended, primarily from universities, and including research administrators, researchers, and professional intermediaries. Others came from government departments, CSIRO or other research bodies.

The Roadshows were designed to provide information about ANDS and its services, and to address issues around the data management aspects of the *Australian Code for the Responsible Conduct of Research*. While ANDS has no responsibility for the Code, it provides a convenient framework for ANDS to discuss data management and sharing and the development of an Australian Research Data Commons.

**eResearch Australasia Workshop and Poster Session:**

ANDS conducted one workshop and a poster session at eResearch Australasia in November, 2010.

“Appraising and archiving data” was conducted with the help of Dr Andrew Wilson of the Queensland State Archives and Gavan McCarthy from the University of Melbourne. About 40 people attended from a wide variety of research institutions and government departments.
“Creating a Culture of Data Citation” was the subject of a poster session designed to publicise the value of publishing data with an assigned DOI which can subsequently be used not just to provide a permanent identifier for the data but to allow the tracking of data citations by the different researchers who use it.

Figure 4: Building a Culture of Data Citation

Targeted training

ANDS Boot Camp

ANDS invited one member of the staff of institutions receiving contract funding through its Seeding the Commons Program to attend an intensive five-day training program, named the “ANDS Boot Camp”. This was intended to equip these staff to take a lead role in their organisation in supporting ANDS funded projects, in passing on their skills and in enhancing their institution’s eResearch support capacity.

The first ANDS Boot Camp was held in early June, 2010. The second and final Boot Camp was held at the Australian National University in Canberra in August 2010. 20 participants, representing either an Australian university or research group attended each 5 day BootCamp.

The program was focused on expanding Australia’s skill base in eResearch across the institution and on participation in the Australian Research Data Commons. The five days covered:

- eResearch, the Australian Research Data Commons, the data curation continuum and institutional models for supporting eResearch
- national frameworks to support eResearch and the Australian Research Data Commons — NCRIS, ANDS, ARCS, NCI, the legal and regulatory environment — as well as policy and legal settings such as Gov 2.0 and intellectual property
- data management and data management planning, metadata for research and metadata stores
- metadata sharing and management
- creating a marketing plan and looking forward

The Boot Camp was well received, with comments including:

- “The bootcamp is a huge networking and building relationship exercise.”
- “well rounded program considering the participants were from different back grounds and were at different stages with ANDS program implementation”
- “I had a great week and learnt a lot. Thanks so much to all the ANDS staff for organising so well. The course folder to put all our notes in each day was very handy.”
- “The Boot Camp was a fantastic event: well organised; well-programmed; great mix of
The practical difficulty of offering training for everyone engaged in projects associated with the Data Capture Program led to the videoing of a workshop for project participants in Melbourne in September 2010. All videos were subsequently edited and made available via the web (http://ands.org.au/guides/data-capture-briefing.html). The workshop was designed to provide a comprehensive overview of aims of the program and to equip participants with the technical basis to undertake their projects.

**Data Capture Briefing**

**ANDS Community Days**

The ANDS Community Days were held in Brisbane, Adelaide, Melbourne, Sydney and Perth. They assisted with the identification of institution-specific issues. This resulted in specific one-to-one discussions and advice. Input from these community days has helped shaped the direction of future service offerings across ANDS. More detailed reporting of this activity is found in the Seeding the Commons activities section.

**Guides and training materials**

Sixteen more guides were made available during the year, and a comprehensive review of all existing guides was undertaken with a view to ensuring currency. The guides fall into two main categories: those supporting data management and those describing ANDS services. One significant set of guides on the topic of data management covers the topic of *Creating a Data Management Framework*. This is designed for institutions who wish to improve their institutional support for data management. It sets out six principles underpinning research data management and identifies four elements within an institutional context to support effective data management. These elements are Policy and Procedures, IT Infrastructure, Services and Metadata Management. The guide is supplemented by additional web pages and a Capability Maturity Guide to assist institutions to assess their own institutional capacity.

Guides describing how to take advantage of ANDS services were supplemented during the year by additional materials on Party Identifiers, Metadata Stores and an updated and expanded ANDS Content Providers Guide.

Most ANDS guides are written by ANDS staff. However, a working level guide to metadata, designed for researchers, was supplied by Intersect.

**Data Commons Policy**

A firm relationship has been established with the ARC, NHMRC and DIISR. A series of meetings has been held at several levels from the Chief Executive Officer to policy officers. The foundation has been laid for a solid program of iterative progress on funding policy and how it relates to data reuse (publishing). This is likely to

30 September 2011

ands.org.au
be achieved via a Pilot Project involving a small number of universities who are ‘early adopters’ in terms of data management and the open access agenda more generally.

A series of meetings have been held with AGIMO as well as senior and chief executives within CSIRO, ABS, GA and the OAIC on the implementation of Gov 2.0. Policy briefs on the implication of Gov 2.0 and the open access agenda more generally for the research sector have been prepared for the ANDS steering committee, ANDS staff, and ANDS training events.

Briefs on the relevance of the Code for the Responsible Code of Conduct of Research have been prepared for ANDS partners, ANDS staff, and ANDS training events. Similar briefs have been developed on the NHMRC deeds and the ARC Funding Rules, which have changed significantly this year.

Model institutional policies for research data management have been drawn up and disseminated through the ANDS community bulletin board

Guides on appraisal policies for research data from an archival viewpoint have been completed and published as ANDS discussion papers. A joint venture is underway with the Digital Curation Centre (JISC) for a shared publication in this area.

ANDS has joined the Commonwealth Spatial Data Management committee and has attended all meetings during this period.

**Licensing Frameworks**

ANDS has played a significant policy advocacy role in transitioning Gilf to AusGOAL, the Australian Governments Open Access and Licensing Framework. AusGOAL now has a ‘research data’ tab and is functionally aligned to the ANDS website. Additionally, ANDS has a place at the governing sub-committee for AusGOAL which is under the auspices of the Cross Jurisdictional CIOs Committee (CJCIOC). To support the significant progress made over the past year in this area, ANDS is planning to create practical licensing working groups in the research and university sectors.

**Analytical Basis**

Very significant progress has been made on a significant report on the value of data. The ABS, GA and the NWC have all contributed to a report entitled *Costs and Benefits of Data Provision* by Professor John Houghton, which is due to be finalised in September 2011. A great deal of interest has been expressed in the report by a range of individuals and organisations, including by the Australian Information Commissioner.

**2.3.4 Program Highlights, Issues and Breakthroughs**

Highlights for this program include:

- A high degree of success in ensuring the skills base required to undertake ANDS projects.
- The firm relationship established with the ARC and NHMRC will lead to a report on how institutions’ data management activities can be functionally linked to the ARC’s funding rules as they relate to data publication in the first instance, and later to the NHMRC’s equivalents.

- We are well advanced in both policy and practical implementation terms (see www.AusGOAL.gov.au) for a common approach to the licensing of publicly-funded data, be they from the government or research sectors.

- ANDS is well advanced on a novel report on the costs and benefits of providing data free of cost, and how this might impact on the individual agency in economic terms; this report has the potential to enable some institutions to move forward in terms of open access, as it identifies and enumerates the costs and benefits in monetary terms.

- The ANDS website and Community Bulletin Board has experienced an impressive growth in “links to” and “links from” over the year as shown in the diagrams below.

<table>
<thead>
<tr>
<th>Ands.org.au (Incoming &amp; Outgoing Links)</th>
<th>October 2010</th>
<th>June 2011</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ands.org.au (incoming &amp; outgoing links)</td>
<td>203</td>
<td>467</td>
<td>230%</td>
</tr>
<tr>
<td>Bulletin board (incoming &amp; outgoing links)</td>
<td>7</td>
<td>39</td>
<td>557%</td>
</tr>
</tbody>
</table>

**Figure 5: Links to and from the ANDS Website and Community Bulletin Board**

2.3.5 **Program Learning**

The capability of researchers, support staff, and research organizations to take advantage of national research infrastructure remains a significant obstacle, and there remains a strong demand for any training, instructions, direction that ANDS can provide. Two areas have been of particular focus in 2010-11:

---

1 For projects commencing in 2012
- the needs of those using ANDS national services
- the needs of our partners institutions as they build out their part of the distributed national infrastructure in the context of ANDS funded projects

In this way 2010-11 has been a year of responding to the needs of people and organizations using ANDS services and doing ANDS funded projects. This has given the program more of an ANDS focus than perhaps first envisaged. This decision has ensured that the ANDS "building capabilities" agenda has been more pragmatic and directly attuned to the immediate needs of our partners and focused on delivering immediately relevant materials to the project context. This is perhaps to the detriment of a broader program of capability-building for data-centric infrastructure at both research organisations and other NCRIS facilities. This remains on the agenda with attention to the infrastructure focus of the NCRIS context.

On the national policy front it is apparent that fundamental adjustments of national policy are a long-term proposition. ANDS should possibly look to broaden its advocacy base by teaming up more closely with national bodies such as CAUL, CAUDIT, Universities Australia, and the scholarly academies to promote particular issues.

### 2.4 Project Office

#### 2.4.1 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.

#### 2.4.2 Outline of projects

The function of this program has evolved as ANDS has matured. Its first major task was to run the consultation process on the ARDC jointly with DIISR. This involved visits to each state to run consultation with all key stakeholders to ensure that the ARDC project met the needs of the stakeholders and had endorsement of the stakeholders.

The next major task was to build a process for the development of ARDC infrastructure by Research Institutions.

The other important activities are to manage contracts – a very large number given the outcomes of the EIF process – and provide appropriate reporting on the ARDC project to DIISR, the Steering Committee, Monash, ANU, CSIRO and public reporting through the web site.

The program has oversight of contract management, business management and communications for ANDS. As the projects progress into varying stages of completion the activity is transitioning from contract establishment to contract monitoring. This will ensure that deliverables are accounted for, milestone payments are made in accordance with the contractual obligations and outcomes are reported.

An aspect of the monitoring will be communicating the outcomes of projects to share the knowledge and deliverables derived from them. This activity will become increasingly important as the number of projects reaching completion increases.
2.4.3 Activity/Deliverables for 2010-11

The majority of contracts were agreed during this period and have now moved into execution stage which requires significant monitoring. A workflow has been designed and implemented in JIRA software to manage this activity.

As the projects reach completion and outcomes are delivered, there is a need to communicate what has been learned and delivered. The program has developed and is now implementing relevant communication strategies to maximise the impact of these investments.

Key reports and business plans have all been produced on time.

2.5 Promotion

ANDS has undertaken a large number of promotional activities during the period July 2010 to the end of June 2011. These include the following.

2.5.1 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:

- Australian Digital Forum, Melbourne
- International Linked Open Data in Libraries Archives and Museums Summit, San Francisco, CA, USA
- Participating in a Metadata Community of Interest, which has been established to influence outcomes for the broader information community and discuss whole of government approaches to metadata issues to enhance the visibility, accessibility and usability of information. This has been an opportunity for ANDS to network with public sector data publishers including ABS, GeoScience Australia, the Office of Spatial Data Management, National Archives of Australia, and AGIMO, the publishers of data.gov.au
- Meta2011 in Canberra, hosted by Institute of Metadata Management
- The IEEE e-Science 2010 Conference, Brisbane
- CAIRSS Community Day 2010, Melbourne
- eResearch Australasia, Gold Coast
- NCCARF Queensland Roadshow, Brisbane
- Metadata Australia Conference 2010, Canberra
- General meeting of CAUL (Council of Australian University Librarians), Queensland
- 15th Australasian Remote Sensing and Photogrammetry Conference (ARSPC), Alice Springs
- JISC (Joint Information Systems Committee) Managing Research Data (MRD) international workshop, Birmingham, UK
- ANDS Roadshow, Townsville
- ANDS Community Events in Brisbane, Adelaide, Melbourne, Perth and Sydney
- ANDS Data Capture Briefing, Melbourne
- ANDS Data Commons Boot Camp, Canberra
- Invitational Research Data Infrastructure Workshop, Prato, Italy (an ANDS-facilitated event)
- Australia-EU Research Infrastructure Workshop, Brussels, Belgium

30 September 2011

ands.org.au
2.5.2 Forums

ANDS has hosted or presented a wide range of forums over the reporting period. The highest profile of these was a series of ANDS Community Events which were run as part of ANDS’ ongoing community building efforts. These were held in Brisbane (QLD), Adelaide (SA), Melbourne (for VIC and TAS), Perth (WA) and Sydney (for NSW and ACT) between February and May 2011, with a total of 193 partner staff in attendance. These partners presented on 92 different ANDS-funded projects.

Feedback from partners indicated that they found hearing about other projects, making connections with others, and learning from other partners and ANDS staff valuable. Having various experts in the room also meant that partners were able to seek clarification on issues that were specifically relevant to their project. During each event, partners broke into discussion groups to either problem-solve or share issues. These discussions have been captured by ANDS to help prioritise future support.

At the end of each Community Event, partners were encouraged to join the ands-partners Google Group and Community Bulletin Board, to help further share information.

In April 2011, ANDS facilitated a three-day invitational Research Data Infrastructure Workshop at the Monash University Prato Centre. The reason for choosing this venue was to facilitate the greatest possible participation from Northern hemisphere invitees. The objective was to bring together research data infrastructure providers and stakeholder representatives in order to explore the potential for much closer collaboration and co-ordinated activity. ANDS will seek to facilitate actions arising from this workshop that are of particular strategic relevance to our next Business Plan, and will encourage action on the other outcomes. ANDS’ staff are already having follow-up meetings to keep the momentum going after the workshop, and starting to deliver on the outcomes. It was agreed by all participants that it had been a successful workshop, with many of them expressing their appreciation to ANDS for bringing this group together.

Other forums included the ANDS Data Commons Boot Camp in August 2010 (for staff working on ANDS-funded Seeding the Commons projects) and the ANDS Roadshow in Townsville in September which provided information on ANDS services and Research Data and the Code for the Responsible Conduct of Research, and was open to all interested parties. In September 2010, ANDS also organized a Data Capture briefing at the Monash Conference Centre for staff of the University of Melbourne, RMIT, Monash University, La Trobe University and the Australian Synchrotron who were engaged in ANDS-funded Data Capture projects. The briefing was designed to provide an introduction to ANDS and its services. A number of participants also provided descriptions of their institutional projects.

ANDS staff, via eResearch SA, have run a series of events during the course of the 2010-11 financial year, called Bright Ideas over Breakfast, at which they have presentations from researchers. To date, all the featured researchers have been involved in ANDS-funded work. Some of the showcased projects have been the University of Adelaide’s genomics data capture project, a number of projects being managed by CSIRO and the Flinders University data capture project centred on Sleepy Lizards.
2.5.3 Consultation meetings

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

2.5.4 Newsletter

The ANDS quarterly newsletter, Share, continues to create awareness of ANDS and its activities amongst the research community and stakeholders by providing updates on ANDS-funded projects, highlighting achievements and promoting ANDS events and objectives. The first 3 issues in the 2010-11 financial year focussed on the new uses of research data. At the start of the 2011 calendar year, the decision was made to have 4 themed issues of share, each focusing on one of the four transformations that ANDS seeks to enable. Thus Issue 8 – the last issue in 2010-11 – released in April 2011, focused on the management of research data, showcasing ANDS-funded projects that are enabling this transformation.

2.5.5 Other activities

ANDS has established a mailing list (ands-general) to promote news about the service, as well as participating in other mailing lists as appropriate. A list for people working on ANDS-funded projects (ands-partners) has also been established, as has a bulletin board - http://community.ands.org.au – to encourage the exchange of information about the creation of the Australian Research Data Commons.

ANDS also hosted a visit from a delegation of Danish librarians interested in ANDS’ data management strategies in October 2010.

2.6 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 June 2011, 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 had been reduced from a rating of high/medium to low. 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 has also been reduced from high to low.
There continues to be medium risk around data providers/federators making their data available and the fact that re-users 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 have been lowered from a high to medium rating. 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. Another risk that has been reduced to a medium rating is that the standards and technologies that ANDS adopts are not adopted more widely. This has been done by participating in review meetings with the providers of these technologies, such as DataCite, ISO2146 and RIF-CS. 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.
## 3 Progress against milestones

### 3.1 Seeding the Commons

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>10Q3</td>
<td>Engagement with and support of funded partners</td>
<td>Achieved: Working with the 8 projects underway as well as engaging with other partners to define potential projects. Records received from 2 projects. Slippage: Fewer records received than expected due to complexity of the task, issues with resources to check records and assist partners, and delays in beginning projects.</td>
</tr>
<tr>
<td></td>
<td>Initial manual feeds of data descriptions available in ARDC</td>
<td></td>
</tr>
<tr>
<td>10Q4</td>
<td>All manual feeds of data available in ARDC</td>
<td>Achieved: Records received from 3 projects. Planning underway for the community events which were held in the first half of 2011. Slippage: Fewer records received than expected due to complexity of the task, issues with resources to check records and assist partners, and delays in beginning projects.</td>
</tr>
<tr>
<td></td>
<td>Community activities identified and commenced</td>
<td></td>
</tr>
<tr>
<td>11Q1</td>
<td>Automatic feeds of data descriptions is underway; Data management planning and policy support delivery</td>
<td>Achieved: Community events assisted in the delivery of data management planning and policy support. Slippage: Fewer records received than expected due to complexity of the task, issues with resources to check records and assist partners, and delays in beginning projects.</td>
</tr>
<tr>
<td>11Q2</td>
<td>Completion of contracted projects as appropriate, including feeds and data management support</td>
<td>Achieved: 2 projects completed. 1 project essentially complete, but extended to deliver more records.</td>
</tr>
</tbody>
</table>
Progress towards goals (including approved records supplied and data management deliverables in place) at 13 projects

Slippage: Not all projects completed, so not all software is available.
Not all projects underway, so not all have commenced testing.

### 3.2 Frameworks and Capabilities

<table>
<thead>
<tr>
<th>Milestone Date</th>
<th>Milestone</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>10Q3</td>
<td>Initial training provided to ANDS project partners</td>
<td>Achieved: The 2 ANDS Boot Camps have ensured most project partners have received initial training. Videos of sessions are available for future training and review.</td>
</tr>
<tr>
<td>10Q4</td>
<td>Workshop planning underway</td>
<td>Achieved: A co-ordinated 2 year workshop plan has been developed and enacted – including internal ANDS training, public events and partner programs.</td>
</tr>
<tr>
<td>11Q4</td>
<td>Scheduled development of support materials guides programs</td>
<td>Achieved: 38 guides and other support documents are now available on the ANDS website. There is a regular review, update and addition cycle in place.</td>
</tr>
<tr>
<td>11Q1</td>
<td>Data management planning and policy support delivery with ANDS outreach programs</td>
<td>Achieved: Creating a Data Management Framework materials are designed for institutions who wish to improve their institutional support for data management. They set out six principles underpinning research data management and identify four elements within an institutional context to support effective data management.</td>
</tr>
</tbody>
</table>
These elements are Policy and Procedures, IT Infrastructure, Services and Metadata Management. The guide is supplemented by additional web pages and a Capability Maturity Guide to assist institutions to assess their own institutional capacity. 


<table>
<thead>
<tr>
<th>11Q2</th>
<th>Planned training/ outreach/ fora/ workshops completed</th>
<th>Achieved: 5 Community Days and were conducted and these were very successful in helping to create communities of interest connecting various ANDS projects. The 9th and final Roadshow was conducted in Sept 2010.</th>
</tr>
</thead>
<tbody>
<tr>
<td>11Q2</td>
<td>Capability building partners identified</td>
<td>Achieved: Partners have been identified through the various outreach activities and this remains a focus area for 2011-2012.</td>
</tr>
</tbody>
</table>

30 September 2011
4 Deviations from the Project Plan

There have been three significant variations from the Annual Business Plan 2010/11 (ABP10/11) but all are consistent with the overall project plan. They are that the proposed approach of developing consortia of research institutions to develop and install metadata store solutions at an organisational level, as envisaged in ABP10/11 was not deemed to be attractive after advice from stakeholders, that the ARDC Applications projects took longer to identify than was envisaged, and that expenditure overall was at a lower rate than was planned for. This section describes all deviations for both projects, as they are jointly planned.

The ARDC Metadata Stores identified the value of having institutionally supported services for both object level metadata and collections level metadata. However ANDS could not fully fund different solutions at every institution, so the approach taken was to fund some exemplar solutions and develop a process to enable institutions to form clusters around solutions that best fit their needs. However the desire to support institutional metadata stores was quite varied and substantial numbers of institutions were not yet in a position to determine their needs, or their preferred solution. At the same time, a number of institutions were progressing their own approaches as part of their ANDS funded Seeding the Commons and Data Capture activities. As a result the proposed process was not adopted and new approaches to metadata stores were devised for execution in the 2011/12 Annual Business Plan.

The ARDC Applications program is intended to demonstrate the value of improved use of research data, and demonstrate this in the light of Data Capture projects whenever possible. Due to the delay in commencement and execution of these projects, there was a consequential delay in the commencement of the Application projects. This delay has had the negative effect that the demonstrations are occurring later than planned, but the positive effect that it is allowing projects to commence that are very often exploiting infrastructure developed by both ANDS and our research data intensive infrastructure partners – notably EMBL, IMOS, ALA, TERN, BPA and others.

Actual expenditure was lower than budgeted. Two of the reasons have just been discussed – slower execution of the Metadata Stores program and the Application Program. The other two reasons were lower expenditure on staff than was budgeted – this was simply the time taken to get good recruits, and the more significant reduction was the time taken to move from a commitment – a letter of offer to an institution, to an agreed project, to an agreed contract to receiving invoices. From a budgeting perspective, we had not modelled this time, assuming that shortly after the letter of offer, there would be a first invoice. This has not been the case, and has actually been extremely beneficial. The pie charts (Fig. 8) show the breakdown of expenditure, commitments, and uncommitted funds showing steady progress in moving commitments to contract. This is further discussed on page Error! Bookmark not defined..

This three year project plan shows an un-even level of expenditure (as ANDS had already made substantial commitments) but does balance the need to engage with the sector over a longer period of time, and to demonstrate value early. The business plan for 2010-11 described the proposed activity over the first year of a three year plan concluding in June 2013, and the following chart shows the intended expenditure pattern for the various programs (showing both NCRIS ANDS and EIF ARDC).
5 Co-Investment

5.1 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.

5.2 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 process of contracting 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.

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.
6 Performance Indicators

6.1 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 20 automatic plus 80 manual metadata feeds. This will cover at least 30 out of the approximately 50 research data-holding institutions that we know about.

   **Result:** 21 research institutions fed collections descriptions to RDA, along with 63 individual collections. From these institutions, 38 individual data source feeds have been set up (13 automatic feeds and 25 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, PFROs, and 4 major Government data providers this year, and through them at least 50 research groups.

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

   - 35 universities
   - Publicly Funded Research Organisations: ANSTO, CSIRO and AIMS
   - Government data providers: ANDS has engaged with over 25 government agencies apart from the PFROs. These include GeoScience Australia (GA), Australian Institute of Health & Welfare (AIHW) and Australian Bureau of Statistics (ABS) directly; Australian Antarctic Division (AAD) and Royal Australian Navy through the engagement with AODN; 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 Depts 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
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: 5

**Result:** 4 – Monash University, University of Melbourne, Queensland University of Technology, Griffith University.

4. The number of times a search is initiated with an ANDS discovery service: 0 – In this year ANDS will concentrate on supporting Google discovery of ANDS data pages, and “see also” services.

**Result:** This indicator is not yet considered important to ANDS current goals, and has not been measured. From September 2011, ANDS will begin to log and measure searches in Research Data Australia and will report these in 2011-12.

5. The number of times an ANDS data page (defined below) is accessed: 100,000 – this is based on a ramp-up as there were few pages that could be discovered at the start of 2009-10, but ANDS will concentrate on directing traffic that way as the pages increase in number.

**Result:** 56,659 filtered page views and 37,527 unique views or visits (from Google Analytics tool). Note: Previously ANDS measured this indicator with raw page hits from the web server; from this year ANDS will report the more standard measure of filtered page view and unique page visits using the standard filters of the Google Analytics tool. Using the old raw web server metrics the result for 2010-11 would have been approximately 201,054 hits (extrapolation from 2011 logs).

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:** Prior to commissioning the survey, ANDS Directors met with senior staff and agreed on the aims and objectives of the survey. The target audience included researchers, research support staff and data custodians.

The aim was to capture understand the community attitude to research data, in particular the areas of focus of ANDS, and then to determine whether ANDS had been successful to date responding to the needs identified by the community survey. This aim recognised that the survey was being undertaken at an early stage in the ANDS program and that not all of ANDS target audience would have had time to comprehend and utilise the ANDS process and support materials.

Insync Surveys was appointed to conduct the Attitudes Survey. Respondents were presented with multiple-choice questions related to managing research data and ANDS. They were also asked to rate their statements in two ways: firstly, to measure the importance of each of the statements to them, and secondly, to measure their impression of ANDS’ performance on each statement.

The survey was completed by 154 respondents who were invited to participate through the various ANDS communication channels. These respondents came from a large pool identified by ANDS staff to cover all identified audiences at sufficient depth to provide significant results.

Insync Surveys reported the following:
Researchers and Research Support Services clients indicated that the most important issues were related to ANDS’ role in improving the management of research data, and helping to publish that data. Other areas included effective communication about ANDS’ online services, and wide consultation with stakeholders.

Among researchers and research support services clients, a number of issues emerged. They recognized ANDS’ role in improving Australia’s capability to manage its research data, but they were less satisfied with ANDS’ communication of its services and support material.

The survey results suggest ANDS did not appear to have consulted across the research sector, and in particular with researchers. The respondents indicated that ANDS could improve on the current performance; this result has a positive aspect in that while ANDS has not met the needs of the stakeholders, they are sufficiently interested to want to hear more and to be engaged in the process.

Clients also indicated that universities and other publicly funded research organizations receive the most attention in relation to ANDS’ promotion of the re-use of data, the development of data management ambitions, the identification of policy issues and the development of data management frameworks.

The survey indicated very substantial differences in the visibility of ANDS amongst different audiences – where highest visibility is with librarians and the lowest with researchers. This accords with ANDS approach of supporting institutions in their engagement with researchers on research data, rather than direct engagement.

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 10 agreements to make data available.

Result: 49 organisations with whom ANDS has agreements and who are creating records in either the public or draft systems.

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 harvestable repositories: 5000

Result: 26,746 collections as at June 30, 2011. 51% of these collections were from Queensland Facility for Advanced Bioinformatics (QFAB) and 31% came from Australian Ocean Data Network (AODN). The number of ANDS research data sets in harvestable repositories increased almost 23 times from last year.

9. The number of research data sets with persistent identifiers: 7000

Result: 6323 persistently identified datasets, 48% increase from last year’s 4279 datasets.

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 soon to be launched data citation service will
encourage publication of data collections with persistent identifiers in a citable form. Consequently we will be able to use international citation services to measure this.

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 there 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.

6.2 Progress over the Life of ANDS Projects

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.

30 September 2011

ands.org.au
<table>
<thead>
<tr>
<th>Measure</th>
<th>2009-10</th>
<th>2010-11</th>
<th>In Progress</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Institutions Participation in RDA (^1)</td>
<td>NA (^5)</td>
<td>21</td>
<td>33</td>
<td>43</td>
</tr>
<tr>
<td>Research Institutions with Data Management Policy and Practice</td>
<td>3</td>
<td>4</td>
<td>19</td>
<td>43</td>
</tr>
<tr>
<td>Institutional context capture tools</td>
<td>0</td>
<td>6</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Institutional Research Metadata Store</td>
<td>0</td>
<td>9</td>
<td>24</td>
<td>43</td>
</tr>
<tr>
<td>Research Data Provider Participation (^2)</td>
<td>NA (^5)</td>
<td>9</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Research Data Infrastructure Partners (^3)</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Data Collections</td>
<td>1,173</td>
<td>26,746</td>
<td>30,000</td>
<td></td>
</tr>
<tr>
<td>Research Data Exploitation Tools</td>
<td>9</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Fields of Research Coverage (^4)</td>
<td>5</td>
<td>21</td>
<td>21</td>
<td>22</td>
</tr>
</tbody>
</table>

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 bar DIVISION 22 PHILOSOPHY AND RELIGIOUS STUDIES are currently covered in RDA

Note 5. Research Data Australia had not been launched as at 30\(^{th}\) June 2010

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

### 6.3 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 the ARDC: This requires that all pieces of infrastructure have been installed to enable a researcher to 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 the ARDC has been established, although 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 Figure 3: Australian Research Data Commons Progress provides this information.

Providing a meeting place for research data management: Measures that indicate research date management plans, and the participation in the ARDC by research institutions, research data providers, and research infrastructure partners all provide an indication that ANDS has achieved this objective. However there are other important indication provided in this report that help demonstrate success against this objective. ANDS has reported increased level of international engagement, and important elements have been triggered by international requests. ANDS staff have been invited to provide international keynote addresses, participation in international forums, membership of DataCite, and notably participation in an EU/US summit on data infrastructure at EU invitation.

Populating the ARDC: ANDS has refined its objectives in this regard: ANDS wishes to populate the commons with collections that are managed, connected, discoverable, and re-useable. 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 lead 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. Many collections are described, with contact details provided that enable the start of a discussion on possible access to the data. Each approach might be highly appropriate, but they are all counted as a collection that is registered and made discoverable through Research Data Australia. Naturally, as reuse increases, which will be able to be tracked through our data citation services, we will be able to determine the value of easily integrated data via web services, compared to mediated access. Consequently it is of value to track each of the properties that we are considering: that data is collected, managed, connected, discovered and used.
7 Appendices

7.1 Confidential Information

There is no confidential information.

7.2 Project Description Detail

The following is a summary of projects agreed, commenced or undertaken during reporting period.

ANSTO
To enhance ANSTO’s scientific data management capabilities for scientific research data, create a base set of metadata for scientific datasets and data collections, and identify, describe and publish metadata on a number of ANSTO’s scientific data collections to the Australian Research Data Commons (Managed by the Australian National Data Service).

Australian National University
Analysis and recalibration of Australian National University’s dataset publication environment to facilitate publication and reuse. Identification of datasets available for publication. Capability building to improve data management at the University.

Central Queensland University
This project seeks to influence researcher attitude and institutional policy at Central Queensland University through an exemplar activity involving curation of data collected by the Centre for Environmental Management. To accomplish this task the development of policies and practical protocols around the management of research data will be required.

The project will be based around longitudinal environmental and social data collected by the Centre for Environmental Management (CEM). This core curation activity will be used to drive policy development and a framework of practical procedures associated with data curation at CQUni.

Thus it is intended that the project will act as an exemplar activity towards the aim of improved research data curation at Central Queensland University.

CSIRO
Seeding the Commons: This project will enable CSIRO’s Biological Collections to be made available in the ARDC.

Curtin University of Technology
This project has the following key aims:
1. To identify, collate and describe research data collections for 20 publications associated with areas of research strength, held by Curtin University of Technology, including those in paper format, and publish high quality metadata to Research Data Australia.

2. To develop and implement university policies and procedures that support the responsible conduct of research, are effective and are based on sector best practice.

3. To develop and implement training programmes for researchers and research support staff with regard to data management.

4. To develop and implement a research data management planning tool that reflects the policies, procedures and best known practice.

5. To develop and implement an ongoing institutional data management framework comprising the policies, procedures, training, tools and support staff.

Deakin University

This project has the following key aims:

1. To record approximately 50 Deakin University research data collections from a number of key research areas on the ANDS Collections Registry to the standards required.

2. To test and develop sustainable processes for the capture of such information into the future, utilising Deakin University’s research repository, Deakin Research Online (a Fez/Fedora system).

Edith Cowan University

This project has the following key aims:

1. To develop university wide data management policies and procedures.

2. To develop and deliver internal training programs and resources for research data management.

3. Identify at least 12 strategically important research data collections at Edith Cowan University associated with publicly funded research and published material (2003 – 2010)

4. Ensure that RIF-CS metadata about these collections are made available to ARDC.

Flinders University

There are two projects at Flinders:

Motion Picture Producers and Distributors of America, Inc. Database

This project will make available a dataset of international significance developed during an ARC Large Grant project entitled Reforming the Movies: A Political History of the American Cinema, 1908-1940, held from 2001 to 2003. The dataset comprises 35,000 digital images (18GB, in JPEG format) of documents digitised from a microfilm copy of the General Correspondence files of the Motion Picture Producers and Distributors of America, Inc. (MPPDA), Hollywood’s industry trade association, covering the period from 1922 to 1939.

The aims of the project are to:

30 September 2011

ands.org.au
- transform the existing MPPDA database to a format that is machine-readable and standards-compliant to support its scholarly use
- store the MPPDA database in a Flinders University research data repository
- make the complete MPPDA dataset, including digital images of archival material, available to authorised scholars from the Flinders University research data repository
- provide a mechanism and process for authorisation of scholarly access to the MPPDA dataset
- describe the MPPDA dataset using RIF-CS Collection, Party, Activity and Service metadata
- store the RIF-CS Collection, Party, Activity and Service metadata on a Flinders University research data metadata store. (The metadata store will interface with all relevant ANDS webservises, including Register My Data, Identify My Data (Pld) and (as appropriate) the pending ANDS Digital Object Identifier (DOI) webservice.)
- develop university-wide data management policies and procedures, using this project and the MPPDA dataset as a pilot
- use this project as a pilot to develop web tools and services to allow users to create, store and manage research data collections along with metadata descriptions (originally in RIF-CS format or mapped to RIF-CS) harvestable by the Australian Research Data Commons, persistent identifiers, and digital object identifiers.

Reformatting the AusStage dataset to support access and re-use by researchers

The aims of the project are to:

- Describe the AusStage dataset using RIF-CS Collection, Party, Activity and Service metadata. Some or all of the RIF-CS metadata may be generated via interface with existing AusStage and/or Flinders University systems.
- Store the RIF-CS metadata on a Flinders University research data metadata store. The metadata store will interface with all relevant ANDS webservises, including Register My Data, Identify My Data (Pld) and (as appropriate) the pending Digital Object Identifier (DOI) webservice.
- Determine the best method to reformat the AusStage dataset into a ‘flattened,’ machine-readable, standards-compliant format to support data re-use, and re-format the dataset as appropriate.
- Develop and document a procedure to encode the AusStage dataset in XML, RDF and/or OWL, and expose the dataset to re-use through an open-access web-based API and/or a common query language such as SPARQL.
- Install an automated mechanism to periodically reformat and export the AusStage dataset into the Flinders University digital research data repository.
- Develop Flinders University policies, processes and systems to facilitate population of the Flinders University digital research data repository with datasets from any eligible researchers, research groups, or research projects, and to facilitate population of the Flinders University research metadata store with descriptions of those datasets.
- Release policy, project documentation, and code into open source via the AusStage website, the Flinders University website, the ANDS website, Google Code, or SourceForge as appropriate.

Note 1: A research data repository and a research data metadata store will be developed as part of the EIF-funded, ANDS Data Capture work at Flinders University (DC23A). It is anticipated that the resulting data repository will house the MPPDA dataset and the resulting metadata store will house the RIF-CS metadata.
Note 2: It is anticipated that development of policies, processes, systems and training around use of the Flinders University research data repository and metadata store will be a combined output of the EIF-funded, ANDS Data Capture work (DC23A) and the two NCRIS-funded Seeding the Commons projects, SC22 and SC22B.

**Griffith University**

Griffith is ready to engage in an undertaking to embed ANDS resources to progress the objective of identifying and describing its research data in preparation for making metadata more accessible to the Griffith research community and the Australian Research Data Service. With this project Griffith intends to identify its datasets created from ARC, NHMRC and Arts Council funded projects, describe these datasets using the RIF-CS Schema using information derived from interviews with researchers, and store these records in Griffith’s research data metadata repository (using Equella). Metadata will be made available to the Australian Research Data Collection from Equella. Data Librarians will be seconded to the project which will be completed by June 2010. This proposal received $125,000 in NCRIS funding from ANDS.

**Griffith University/Macquarie University**

Establishment of an Australian National Corpus that:

- Aggregates data from existing corpora residing at Australian universities, to provide a diverse and accurate representation of written and spoken languages in Australia.
- Allows the discovery, access and deposition of written and spoken data
- Allows linguistic researchers to collaboratively apply textual annotations to written and spoken data
- Is multimodal: supports text, multimodal text, audio and AV
- Is multilingual: English in Australia, Indigenous languages, community languages and sign languages
- Increases the potential for re-use of linguistic research datasets, and enables new research opportunities
- Is harvestable by the Australian Research Data Commons (ARDC)

**James Cook University**

This project will identify and create collection descriptions for a wide range of datasets including:

- QLD Government data, such as the Light Detection And Ranging (LIDAR) datasets,
- The Reef and Rainforest Research Centre (RRRC),
- The Centre for Tropical Biodiversity and Climate Change (CTBCC), and
- The Australian Centre for Tropical Freshwater Research (ACTFR).

In addition, the project will identify, describe and manage other significant data collections related to Tropical Data at James Cook University.

These collections will be made available to researchers via the JCU Tropical Data Hub and will be harvestable via Research Data Australia.
La Trobe University
This project will build the platform for a potential national database of historical archaeological collections, excavated sites and the people connected to those objects and places.

Each year archaeologists (both academic and private consultants) excavate tens of thousands of artefacts from historical archaeological sites across Australia. While some states (e.g. Victoria) require catalogues to be prepared in a standard format, the majority of catalogue data are stored in small, standalone spreadsheets or custom-built databases, and few are made freely available. There is no central register of these individual datasets, and many significant collections are simply unknown to archaeological researchers.

Between 2001 and 2004, the La Trobe led ARC-Linkage project "Exploring the Archaeology of the Modern City (EAMC) created two research databases" (www.latrobe.edu.au/amc/) which offered, for the first time, a central database of 700,000 artefacts from multiple historical archaeological sites and a companion dataset of historical occupancy data relating to 2200 individuals who occupied those sites.

While extensive in their data content, the databases themselves are too limited in their structure and require significant design input to make them truly effective tools for managing and sharing historical archaeological data.

This project will contribute to the Seeding the Commons program by federating two EAMC databases together. Conducting data auditing for existing data sets; and seeking new datasets from the private, public and tertiary sectors. It will enable researchers to access a vast dataset that is currently unavailable to them, and provide the platform for future datasets to be made freely available in a standardised and timely fashion.

Macquarie University
This project comprises multiple related activities to advance the process of data management and meta-data collection for research datasets. These activities are: the implementation of a metadata store; the support of data management practices in the Faculty Business & Economics; an audit of the research datasets in the Faculty of Business & Economics and the Faculty of Arts; and a review into the means of expanding the learning and processes implemented in the project to supporting data management across the whole of Macquarie University.

Monash University
Monash University Library will contribute to ANDS objectives by employing staff (2 FTE) on a project to systematically: identify data collections from completed publicly-funded research projects; describe these collections in ways that will facilitate the harvesting of RIF-CS-compliant metadata for ARDC; and improve the storage, discoverability and accessibility of the datasets.

This process will be augmented by ongoing efforts to build workflows relating to new research projects. As with all of Monash University Library’s research data activities, knowledge transfer to other institutions through ANDS forums and events will be a priority.
This project builds on existing activities, such as those led by the Library’s Data Management Coordinator and ARROW Librarian, and the ongoing work in building data management expertise in the contact librarians through the Library’s Dare program.

**Murdoch University**

This project has the following key aims:

- To integrate data from the wheat genome with environmental data (forming a new research data product), which will help devise strategies to increase yields on diminishing tracts of land.

  Data to be integrated to be attained from numerous parties, these include:
  
  a. Data logged by wheat growers and their farming equipment. The datasets will cover:
     - Geospatial
     - Vegetation
     - Wind
     - Soil Moisture
     - Crop temperature and rainfall
     - Soil structure
  
  b. Data sourced from state and national government departments relevant to Australian agriculture. These datasets will cover:
     - Patched-point climate (rainfall, temperature, evaporation) data source – Queensland Department of Natural Resources and Mines
  
  c. Data sourced from CMap wheat genome database. This is curated, updated and maintained by the Centre for Comparative Genomics (CCG).

- To create a web based application, which allows the on-going submission and access of data collected by growers.

**QUT**

QUT is ready to engage in an undertaking to embed ANDS resources to progress the objective of identifying and describing its research datasets in preparation for making metadata more accessible to QUT research community and the Australian Research Data Service. With this project QUT intends to identify its datasets created from ARC, NHMRC and Arts Council funded projects, describe these datasets using the RIF-CS Schema using information derived from data interviews with researchers, and store these records in QUT’s research data metadata repository (using Mediaflux). Metadata will be made available to the Australian Research Data Collection from Mediaflux. Data Librarians will be seconded to the project which will be completed by June 2010. This project received $125,000 in NCRIS funding from ANDS.

**RMIT University**

The objective of this project is to create a Screen Media Research Archive by focusing on two research repositories held at RMIT and on currently held and future research data that RMIT is the custodian of in this research field. The project aims to lift the research curation practice in this specific field, support the researchers in this curation, simplify and partly automate - by organisation, process and software development - this process of curation for years to come. As part of this development, a large number of
currently existing data collections will also be curated to improve the competitiveness and impact of this research group and as a demonstrator for the project.

This project aims to:

- Integrate existing Australian screen media research infrastructure and content residing at RMIT
- Design and develop detailed metadata schema and data management protocols that will establish a national management standard for the digital archiving of higher education screen media research in a variety of formats and for the purpose of the diffusion of Creative Works
- Make available information about screen media research, and where permitted, the research works themselves as full digital assets.
- Enhance and support research and learning on Australian screen media and by Australian screen media researchers.
- Use new technologies to assist discovery.
- Ensure the sustainability of the Screen Media Research Archive through ongoing administration by the AFI Research Collection and development of appropriate metadata descriptions for format migration timelines that provide alerts when assets need to be transferred to a new format.

Swinburne University of Technology

The overall aim of the project is to stimulate the development of a research data management culture at Swinburne. This has a compliance aspect in meeting requirements such as The Code, but also contributes to the movement towards improved public access to the full range of research outputs in Australia.

The project aims to achieve this by:

- establishing user friendly tools and workflows required to facilitate the efficient capture and provision of research data information
- extending existing technical systems where possible to minimise change and development requirements, and to ensure sustainable supported outcomes
- utilising existing staffing where possible to embed data management knowledge in the local culture, and ensure continued activity beyond the life of the project
- providing a number of practical and relevant examples of effective research data management at Swinburne

University of Adelaide

Research data interviews with University of Adelaide researchers. Gather requirements to aid development of specifications for pending University of Adelaide research data repository and metadata store. Development of information and guidelines to support research data management policy.

University of Canberra

Develop research data repository and metadata store. Integrate research data metadata store with other relevant University of Canberra systems, including research grant and contract management systems. Begin development of a research data management policy.
University of Melbourne

This project has the following key aims:

- To identify existing important research collections at the University of Melbourne (UoM). Selected data collections will be described using RIF-CS. The descriptions will be stored on site in a metadata store (Vitro) and harvested into ARDC.
- To develop on-going processes to capture research data with its associated metadata and contribute RIF-CS metadata to ARDC via an OAI-PMH harvest point.
- To develop university wide data management policies and local discipline based procedures.
- To develop and deliver internal training programs and resources for research data management.
- To create web software tools and services to allow users to create, store and manage research data collections along with their metadata descriptions, persistent identifiers and digital object identifiers.

University of New England

The N. C. W. Beadle Herbarium will have its collections identified, described and registered with the Atlas of Living Australia and Research Data Australia. This project is a pilot that will be used to further expand data management and description of other research data collections at the University of New England.

University of New South Wales

Expose specified university data collection metadata to ARDC, internal data management policies, staff and researcher training programs, tools and services for UNSW research data/metadata collection, tools for research data links to publications

University of Newcastle

The Newcastle Research Data Online Project will enable the university to contribute to the Australian Research Data Commons (ARDC). The project will implement strategic systems, interfaces and processes to enable capture of metadata for our Research Data Collections. We will work collaboratively with the University of Southern Queensland who are developing the software and systems that we will implement. In addition, we will work collaboratively with Swinburne University of Technology who will also be testing and implementing the system at their site.

The project will also identify, develop and implement interfaces or linkages to pre-determined strategic triggers within the organisation which will assist with the identification of potential relevant metadata capture, e.g. research storage, self-identify GUI, central registry, Grants database, etc.

University of Queensland

The UQ Seeding the Commons project will have the following objectives:

- To identify UQ-based data collections resulting from publicly funded projects that can potentially be shared via the Australian Research Data Commons;
- To identify any potential barriers to sharing publicly funded data collections and develop strategies and/or mechanisms to overcome these barriers
- To establish a UQ Data Collections Registry that:
a. Enables the registration of new collections and streamlined generation of metadata/collection descriptions that include metadata about the researchers/parties and research projects/activities associated with each collection
b. Provides a search and browse interface to UQ data collections via RIF-CS compliant descriptions
c. Is linked to the Australian Research Data Commons (ARDC) via an OAI-PMH interface and Atom/RSS feed

- To increase the number of research datasets that are available for re-use both within UQ and nationally
- To increase the uptake of data stores and related interfaces, provided by UQ and/or external providers (e.g. ANDS, ARCS).
- To document the methodologies employed in the development of the UQ Data Collections Registry and inform national and institutional policy development, decision-making and project planning
- To evaluate the architecture and processes established for the project in terms of longer term maintenance of an ongoing operational process.
- To expand local expertise and capabilities in data management
- To share any software, data models, metadata schemas or other technologies that emerge from the UQ Seeding the Commons or other ANDS-related activities with the national community
- To develop data management resources for UQ that can be used by other organisations in the ANDS community, including:
  a. A UQ Policy on research data management;
  b. Strategic plans and technical frameworks for both current and future research data storage and management within UQ
  c. Governance structures for data management, including membership and terms of reference for groups;
  d. Outreach and training materials for inductions, workshops and information sessions targeted at research staff, Higher Degree by Research (HDR) students and data management support staff;

**University of South Australia**

The project aims to:

- Make Architecture Museum (AM) collection metadata widely available
- Create sustainable processes to transfer the Architecture Museum collection metadata to an institutional repository.
- Create software processes to expose AM collection metadata to the ARDC and thereby make the associated data visible, shareable and reusable.

**University of Sydney**

The Seeding the Commons activity at the University of Sydney will contribute to the establishment of data management planning and implementation processes, integrated with research practice, that will support storage, identification, discovery and, where appropriate, sharing of primary research data. The outputs of the project include: data management planning guidelines and support materials; the identification and description of significant datasets and the development of procedures and practices to support researchers in the management of research data.
University of Tasmania

Describe the entire contents of the TPAC digital library using RIFCS. ID and serve other research collections held at UTAS. Discover IMAS/TAFI marine community data - descriptions stored in IMAS geonetwork. Develop and deliver internal training programs and resources for research data.

University of Technology Sydney

The UTS Library, ITD and RIO will collaborate to develop an effective process and relevant tools to enable the capture, storage, access and reuse of data and metadata created at UTS.

Building on the existing work being undertaken at UTS for the NSW Node of Australian Social Sciences Data Archive (ASSDA) and the national archive for Aboriginal and Torres Strait Islander materials (ATSIDA), the project will identify ARC and NHMRC funded research.

The project will include:

- Identification of relevant national and international data archives for all disciplines (eg ASSDA for the Social Sciences)
- Identification of disciplines without national or international data archives and a strategy for storage of this data
- Development of process maps for the ingest of data into national and international data archives
- Development of checklists, guidelines and other tools to ensure simple, effective data management planning
- A strategy for communication and promotion of effective data management planning
- Protocols and strategies for archival practices, including reuse and access permissions
- Strategies for exposure of relevant metadata across disciplines

University of Western Australia

The aim of the project is to develop and implement a framework for storage of metadata about research data collections. The specific objectives will include:

- Develop an informed policy framework for institutional research data management
- Educate research staff and students in informed research data management
- Identify at least 50 strategically important research data collections at the University of Western Australia associated with publicly funded research and published material (2003 – 2010)
- Ensure that RIF-CS metadata about these collections are made available to ARDC

University of Wollongong

This project has the following key aims:

- To identify research data collections at the University of Wollongong. The selected data collections will be described using RIF-CS. The descriptions will be stored on site in a metadata store and harvested into ARDC.
- To develop university wide data management procedures relevant to UOW.
Capture metadata for at least 50% of identified research data generated from Australian Competitive Grants from 2005-2010 where datasets exist. There are 404 successful ACGs with potential data collections. Part of this project is to find where datasets exist.

Victoria University

The project aim is to establish and trial a framework for storage of research data. Specific objectives will include:

- Capture at least 50% of research data associated with publicly funded research and published material (2003 – 2010) and make metadata available to ARDC
- Develop on-going processes to capture research data including associated metadata available to ARDC
- Develop an informed policy framework and institutional research data management plan
- Educate research staff and students by informed research data management

### 7.3 Progress against activities

#### 7.3.1 Frameworks and Capabilities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influencing Government policy on research data through papers and discussion, including on Gov2.0 activities</td>
<td>Discussions held regularly with ARC, NH&amp;MRC and DIISR. Submission sent to the Office of the Australian Information Commissioner on Information Policy. Meeting with Information Commissioner and FOI Commissioner.</td>
</tr>
<tr>
<td>Enabling improved data provision to research through influencing policy and procedure, including licensing approaches and determining the economic value of sharing data</td>
<td>Membership of AusGOAL committee. Cost-benefit study contracted to Professor John Houghton of VUT (to be delivered 2011-12).</td>
</tr>
<tr>
<td>Strengthening research data requirements, including potential amendments to funding agreements and the Australian Code for the Responsible Conduct of Research</td>
<td>Discussions with funding bodies have been held and a project has been proposed by ANDS to pilot data plans in funding applications. Early adopter universities have been identified and have had input into the proposal. Further discussions with funding bodies are scheduled. Significant interaction with the EU research funding bodies has taken place and ANDS is actively pursuing integration and</td>
</tr>
</tbody>
</table>

30 September 2011
<table>
<thead>
<tr>
<th>Activity</th>
<th>Progress</th>
</tr>
</thead>
</table>
| Providing approaches that support best data management practice and re-use by researchers and groups. | New guides have been published: New guides  
- *AusGOAL* [PDF]  
- *Creating a Data Management Framework* [PDF]  
  - Institutional policies and procedures  
  - IT infrastructure  
  - Research Data Management Framework: Capability Maturity Guide [PDF]  
- *Metadata (Working Level)* [PDF] Updated July 2011  
- *Metadata Stores Solutions (Working level)* [PDF] Updated July 2011  
- *What is Research Data?*  
Updated guides  
- *Data management planning (Awareness level)* [PDF]  
- *File formats (Awareness level)* [PDF]  
- *File formats (Working level)* [PDF]  
- *Metadata (Awareness level)* [PDF]  
- *Storage* [PDF]  
- *The Data Curation Continuum* [PDF]  
- *Data citation* [PDF]  
ANDS Content Providers Guide established to |
<table>
<thead>
<tr>
<th>Activity</th>
<th>Progress</th>
</tr>
</thead>
</table>
| Develop a shared base of resources used by the community of practice research support staff | Second ANDS Boot Camp held August 2010 as networking opportunity.  
ANDS Community Bulletin Board  
[http://community.ands.org.au](http://community.ands.org.au) and Partners  
Google Group set up to store documents and exchange ideas and issues. |
| Engage with the ANDS Community to support the development and implementation of institutional research data management policies | [What is Research Data?](http://community.ands.org.au) developed in consultation with community.  
One-on-one consultation with several institutions who have sought help with policy development. |
| Train staff working on ANDS projects in the fundamentals of data and information management | Second ANDS Boot Camp held August 2010. All training materials made available via the ANDS Community Bulletin Board.  
Workshop held eResearch Australasia November 2010 on selection of data for long term curation.  
ANDS Content Providers Guide established to support data collection description registration and publication  
Data Capture Briefing held September 2011 captured on video for the use of others  
| Continue ANDS Roadshow series                                            | Completed September 2011, so that all capital cities plus Newcastle and Cairns were included.                                                                                                        |
| Develop targeted documentation to support the improvement of data management, citation, analysis, curation and preservation | All ANDS guides updated and more created to cover these topics. See list below.  
ANDS Content Providers Guide established to support data collection description registration and publication  
New guides as listed above  
Poster session at eResearch Australasia 2010 on creating a culture of data citation. Poster |
Activity | Progress
--- | ---
Provide access to documentation through various media and through direct contact to support widespread access and use and reuse of research data | Extensive updating of ANDS website and creation of new materials to support data sharing and reuse. Video provided of Data Capture Workshop.
Contribute to the organisation of workshops and seminars sponsored by other ANDS Programs | Attendance at presenting to all ANDS Community Days held Sydney, Brisbane, Melbourne, Adelaide, Perth.
Enabling improved data provision to research through determining the economic benefits of sharing data | Project underway due to complete in early 2011-12
Provide select organisations with resources to undertake capability building activities | Second ANDS Boot Camp held August 2010 involved 20 staff of partners engaged in ANDS contracts.
Develop capability through workshops, seminars and other forums. The process for doing this involves identifying appropriate opportunities and then seeking the approval of ANDS management | Second ANDS Boot Camp held August 2010. Workshop at eResearch Australasia November 2010. Data Capture Briefing September 2010.

7.3.2 Seeding the Commons

**Activity**

To grow the data commons, ANDS will emphasise the recruitment of existing content into repositories, identifying existing repositories of useful content, and making all that content discoverable through the ARDC.

**Progress**

ANDS has encouraged this primarily through the EoI projects. The projects have generally taken one of three forms:

- **Broad**: a wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy (21 projects take this form)
- **Exemplar**: working with exemplar data collections within the institution, with a view to applying the lessons learnt to other areas, and to the broader data management policy framework within the institution (10 projects take this form)
- **Combined**: working with the Data
| Where institutions with valuable existing content do not have the required systems, ANDS will work with them to improve their ability to store, describe, persistently identify and register their research data assets, in collaboration with the Metadata Stores program. | Capture projects funded by ANDS, to apply the lessons learnt and help create a broader data management policy framework (2 projects take this form)

The funding of the National Linguistics Corpus project is also designed to make more useful content available through the ARDC. |
|---|---|
| Where repositories (or federations) already exist, ANDS will assist with their integration into the ARDC. | ANDS staff are closely involved in this through the EoI projects. Common elements of these projects include:

- Identification of existing data sets within an institution, with a view to enabling the exposure and description of them were possible
- Examining existing processes, policies and workflows, in data management with the intention of using what is learnt to improve data management capabilities within the institution
- Creating descriptions of datasets to increase understanding of how these descriptions can best be created
- Building tools to enable and simplify metadata capture related to datasets

Regular interaction with the Metadata Stores program occurs to provide advice. |
| To focus funding effectively a large number of Expressions of Interest have been sought to understand better the needs and aspirations of institutions in this area. This will provide funding for institutions to better understand their data management needs, their current data holdings, and put in place systems to manage and exploit this data. | Being undertaken as part of the funding program, at an institutional level.

The funding of the National Linguistics Corpus project is also designed to make more useful content available through the ARDC. |
| At June 30, 2011, ANDS had either entered into contracts (or had substantially agreed on project descriptions) for Seeding the Commons projects at all of the EOI institutions with the exception of the University of Western Sydney and the University of Southern Queensland. A breakdown of the progress made in relation to this is provided in section 2.2.3. Detailed descriptions of the contracted and agreed |
ANDS will also aim to share lessons learned and examples of best practice across the sector to create a community of data managers. This will involve bringing together data managers either directly employed by ANDS, or funded by ANDS, from across the country, to effectively share knowledge and experiences.

To achieve this, ANDS will need to develop an effective set of internal resources to centralise the necessary information and then to disseminate it widely.

ANDS will also work with other programs, key researchers, local bodies and overseas institutions to identify tools and infrastructure that could be co-developed to improve the quantity and quality of the data that is managed, and increase the richness of the contextual information around the data that is available.

Seeding the Commons have worked with staff in other ANDS programs (notably ARDC Core, Frameworks and Capabilities and Public Sector Data) to improved data management service offerings.

The National Linguistics Corpus project is a collaboration between Griffith and Macquarie Universities, and has the support of the Australian Academy of Humanities, Australian Linguistics Society, the ARC Network in Human Communication Science and a consortium of Australian universities. Further work on this area is expected to occur in conjunction with RDSI.

Key researchers have been involved in the development of tools through the EoI projects.

Overseas engagements have largely fallen outside the scope of this program, and been pursued in other programs.
7.4 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.

7.4.1 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

Risk Mitigations:

- Update the communications plans to ensure that the specific e-Research communities have visibility of specific projects and their outcomes before, during and after the projects are undertaken. New diagnostic strategies have been implemented and run to mitigate against failure.
- Provide a central point where progress towards the Australian Research Data Commons can be tracked by metrics such as number of collections available, and numbers of datasets accessed
- Clearly articulate the Project’s message and brand
- Engage carefully 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 DIISR officers to ensure they provide input to decision making, including having an observer on the Steering Committee

Risk 2 – That the Project is not managed effectively

Risk Factors:

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

Risk Mitigations:

- Put in place management and planning processes 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 are in place to facilitate working together.
- Negotiate with DIISR on appropriate administrative staffing for EIF funded programs.
- Staffing levels are monitored and adjusted as required.
Contracts and partnerships with state based organisations that host Project staff will be 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.

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:
- Seek approval for stream-lined approaches at Monash University to enable ANDS to work more effectively.
- Fund additional staff or specific work at Monash University to meet the requirements for bursts of activity.

7.4.2 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.

Risk Mitigations:
- Maximise the effectiveness of connections between the Project and related PFC and other initiatives, including involvement of groups outside ANDS in the ANDS Policy Forum, the ANDS Technical Forum, and the ANDS Content Forum.
- Ensure continuing wide consultation following the consultation on the Draft Final Australian Research Data Commons Project Plan.
- 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 current and foreshadowed NCRIS capabilities.
- All activity plans should be highly inclusive of relevant 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.

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.
- Put in place effective vendor and partner engagement approaches.
- Implement formal procurement processes 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 seeks approval to expend existing funding over longer timelines (consistent with other Superscience-funded activities).
- ANDS Steering Committee seeks additional funding for 2011-12 and 2012-13.
- ANDS creates reliable sustainable services that are offered over the longer term by other long term service providers.
- Strong contribution to DIISR Roadmap process will be a mitigating factor.

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

Risk Factors:

- ANDS and the offerings of other eResearch providers are confused by possible users.
- Relationship between ANDS and MARCS (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.

Advocate for greater coordination of offerings by e-Research service providers through eResearch Infrastructure.

Discussion with NCI, NeCTAR and RDSI taking place to ensure clarity of eResearch service offerings.

7.4.3 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.
- Researchers do not trust the Project’s data sharing and access control mechanisms.
- Researchers are working with other collaborators who have confidentiality concerns over the data.
- 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.
- Recognise researchers through peer feedback for the deposit of data into the ARDC via increased citation – would need to be recorded and measured as a performance measure by the Project.
- Effective communication of structures in place to ensure building of trust.
- Recommend that funding be linked to the provision of data via the ARDC as it becomes available.
- Provide targeted assistance to data federations to assist with integration into the ARDC.

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

Risk Factors:

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

Risk Mitigations:

- Ensure a nuanced and multi-faceted approach to exposing the Project’s accessible data.
- Work with the Australian Access Federation to identify a simple set of standard access control policies.
- Ensure that it is easy to re-purpose ARDC accessible data.
7.4.4 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.
- People with second order skills end up being employed because of staff shortages.
- Limited tenure roles potentially on offer within the Project are not attractive to candidates.
- Sustained high workload leads to staff burnout within the ANDS Management Team.

Risk Mitigations:

- Commence recruitment early to mitigate delays in the commencement of activities.
- Be highly selective in recruitment and favour quality of candidates over the quantity of candidates (do not fill jobs for the sake of it).
- Encourage secondment of staff at an institutional level.
- Investigate non-traditional sources of potential staff.
- Manage staff time and monitor levels of work.
- Engage in project based activities where more leadership roles are taken by senior ANDS staff who are not management.