Australian National Data Service (ANDS)

BUSINESS PLAN 2013-14
# INDEX

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Executive Summary</td>
<td>3</td>
</tr>
<tr>
<td>2 ANDS Context and Approach</td>
<td>4</td>
</tr>
<tr>
<td>3 Status of Project</td>
<td>10</td>
</tr>
<tr>
<td>3.1 Establishment</td>
<td>10</td>
</tr>
<tr>
<td>3.2 NCRIS funding</td>
<td>11</td>
</tr>
<tr>
<td>3.3 Super Science Funding</td>
<td>11</td>
</tr>
<tr>
<td>3.4 Additional Funding</td>
<td>11</td>
</tr>
<tr>
<td>3.5 Time Extension and Reporting</td>
<td>12</td>
</tr>
<tr>
<td>3.6 The Function of ANDS</td>
<td>13</td>
</tr>
<tr>
<td>3.7 ANDS Programs and their Transformation</td>
<td>13</td>
</tr>
<tr>
<td>3.8 ANDS Principles</td>
<td>14</td>
</tr>
<tr>
<td>3.9 Scope</td>
<td>15</td>
</tr>
<tr>
<td>3.10 ANDS’ Impact</td>
<td>16</td>
</tr>
<tr>
<td>4 Research Infrastructure</td>
<td>21</td>
</tr>
<tr>
<td>4.1 National Collections</td>
<td>22</td>
</tr>
<tr>
<td>4.2 National Services</td>
<td>31</td>
</tr>
<tr>
<td>4.3 Institutional Engagement</td>
<td>42</td>
</tr>
<tr>
<td>4.4 International Collaboration</td>
<td>77</td>
</tr>
<tr>
<td>4.5 Overall 2013-14 Outcomes</td>
<td>81</td>
</tr>
<tr>
<td>5 Program Engagement Strategies</td>
<td>83</td>
</tr>
<tr>
<td>6 Promotion</td>
<td>84</td>
</tr>
<tr>
<td>7 Access and Pricing</td>
<td>86</td>
</tr>
<tr>
<td>8 Governance, Management and Implementation</td>
<td>86</td>
</tr>
<tr>
<td>8.1 Governance Framework</td>
<td>87</td>
</tr>
<tr>
<td>8.2 Risk Management</td>
<td>89</td>
</tr>
<tr>
<td>8.3 Milestones for 2013-14</td>
<td>95</td>
</tr>
<tr>
<td>8.4 Key Performance Indicators</td>
<td>98</td>
</tr>
</tbody>
</table>
1 Executive Summary

The Australian National Data Service (ANDS) was established in January 2009 following the ANDS Establishment Project. ANDS was originally created as part of the National Collaborative Research Infrastructure Strategy (NCRIS) initiative to ensure that research data is used as effectively as possible by Australian researchers. The Super Science initiative announced in May 2009 provided additional funding from the Education Investment Fund (EIF) to establish the Australian Research Data Commons (ARDC). This provided the opportunity to leverage both NCRIS and EIF funds to build a co-ordinated set of programs through to June 2013. As a result of additional Collaborative Research Investment Scheme funding, and an agreed reduction in expenditure in 2012-13, ANDS has been funded to operate until December 2014.

ANDS exists to transform Australia’s research data environment by making Australian research data collections more valuable though managing, connecting, enabling discovery and supporting the multiple use of this data. The purpose of this activity is to enable richer research, more accountable research, more efficient use of research data, and improved provision of data to support policy development. The outcome of this activity will be that Australia’s research data as a whole become a nationally strategic resource.

In this document ANDS’ plans for the 2013-14 financial year are described in detail. This plan is in accordance with our extended strategic direction and describes a new model of working where services and expertise are provided to our partners instead of funding. Rather than institutions just focussing on meeting ANDS’ goals, the team will encourage organisations, research groups and researchers to realise their research data ambitions.

There are significant consequential changes to the way ANDS will organise its work in 2013-14 compared to 2012-13 so that it is possible to easily transition to CRIS funding late in the business plan period. The work of ANDS will be accordingly be carried out under four programs:

- **National Collections** of government, institutional, discipline and national facility data
- **National Services** including registration, publication, discovery, and advisory services
- **Institutional Engagement** with all of Australia’s major research organisations
- **International Collaboration** with data infrastructure providers to ensure that Australian research data infrastructure is compatible with international approaches.

Significant progress has been made to date and by June 2013 ANDS will have:

- established several national services: a data collections registration service, a dataset identification service through the DataCite consortium, a data collection description publication service, a researcher identification service in partnership with the National Library of Australia, and Research Data Australia – a data collections discovery service
- populated the ARDC with data collections that have been managed, connected, and have 55,000 collections descriptions discoverable through Research Data Australia, Google and other mechanisms
helped establish coherent institutional research data infrastructure at 32 Universities, CSIRO, the Australian Synchrotron and ANSTO, including automated tools for capturing rich metadata from instruments, metadata stores, and pipes connecting to institutional systems and national services

improved the ability of the Australian research system to exploit the ARDC with guidance on research data management, including responding to the *Australian Code for the Responsible Conduct of Research*

improved data management around the country with a substantially increased cohort of research data managers

developed new tools that enable more effective re-use of research data, with this use demonstrated through several high profile examples.

During 2013-14, ANDS will:

- establish new national collections in identified priority areas that enable new questions to be asked addressing nationally significant research challenges
- support more institutions in managing their research data as an institutional asset
- enhance Australia’s leading role in research data infrastructure internationally
- maintain ANDS as a trusted partner of research institutions beyond funding projects
- continue to deliver valued and reliable national data services
- populate the ARDC with more data collections that have been managed and connected, ensuring over 60,000 collections descriptions discoverable through Research Data Australia, Google and other mechanisms.

By the end of 2013-14 it is expected that researchers across every discipline and at nearly every research institution will be represented in the Australian Research Data Commons, and nearly all research institutions will have improved their research data management, leading to routine publication of their data with ANDS persistent identifiers into a data store that feeds information to the ANDS collections registry. In addition, researchers will be able to find a wide variety of data sets using the ANDS data pages through a variety of discovery paths, and more institutions will be successfully engaged in meeting their responsibilities described in the *Australian Code for the Responsible Conduct of Research*. ANDS will have helped Australia maintain a strong international role in research data infrastructure. Importantly ANDS will be a trusted partner of the research community in transforming research data practice.

## 2 ANDS Context and Approach

Research is becoming more data intensive, and the data is becoming more complex. Moreover the problems being tackled are increasingly large scale and span multiple disciplines. Consequently, and as a result of high level Government reviews, the Department of Innovation, Industry, Science, Climate Change, Research and Tertiary Education (DIISCCRTE) has invested in improving the
Australia’s research sector’s capability to use and re-use research data. This was guided first by the NCRIS roadmaps and then by the document entitled *Towards the Australian Data Commons*\(^1\) (TADC).

In support of this goal, *Towards the Australian Data Commons* identified a range of objectives for ANDS. These objectives were based on the belief that “ANDS can contribute most effectively by developing services and activities that enable stewardship within multiple federations of data management and data user communities” (p. 6). TADC identified a number of longer-term objectives for data management:

a) A national data management environment exists in which Australia’s research data reside in a cohesive network of research repositories within an Australian ‘data commons’.

b) Australian researchers and research data managers are ‘best of breed’ in creating, managing, and sharing research data under well-formed and maintained data management policies.

c) Significantly more Australian research data is routinely deposited into stable, accessible and sustainable data management and preservation environments.

d) Significantly more people have relevant expertise in data management across research communities and research managing institutions.

e) Researchers can find and access any relevant data in the Australian ‘data commons’.

f) Australian researchers are able to discover, exchange, reuse and combine data from other researchers and other domains within their own research in new ways.

g) Australia is able to share data easily and seamlessly to support international and nationally distributed multidisciplinary research teams. (p. 6)

As a result of these goals, initial activity, and consultations, ANDS role is to enable Australia’s research data to be transformed:

<table>
<thead>
<tr>
<th>From Data that are:</th>
<th>To Structured Collections that are:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Unmanaged</td>
<td>• Managed</td>
</tr>
<tr>
<td>• Connected</td>
<td>• Connected</td>
</tr>
<tr>
<td>• Invisible</td>
<td>• Findable</td>
</tr>
<tr>
<td>• Single use</td>
<td>• Reusable</td>
</tr>
</tbody>
</table>

This will deliver a nationally significant resource so that Australian researchers can easily publish, discover, access and use Australian research data.

ANDS is doing this by creating the Australian Research Data Commons (ARDC), the focus of the Super Science project. 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](image)

**Figure 1: Australian Research Data Commons**

ANDS is thus creating a combination of national services and coherent institutional research data infrastructure, combined with the ability to exploit that infrastructure with tools, policy and capability. To deliver against these objectives, ANDS has four inter-related programs of activity (Institutional Engagement, National Services, National Collections, and International Collaboration).

The ANDS activities (whether NCRIS funded or EIF funded) are both being conducted under the same management structure which can thus maximise opportunities for cost savings. However, as required by contract, separate reports will be provided to DIISCCRTE for the NCRIS service and EIF project. ANDS has also entered into four other Funding Agreements with the Department. The first agreement was to progress outcomes arising from the First Joint Australia-EU Research Infrastructure (RI) Workshop in June 2011. Subsequently, in 2012, two more agreements were
entered into, for the Second Joint RI Workshop, and the DataWeb Forum. The most recent funding agreement was to maintain this research infrastructure under the Collaborative Research Infrastructure Scheme (CRIS). This business plan provides a combined view of how ANDS intends to execute all of these activities in 2013-14.

The ARDC Project represents a very significant change in the approach to research data in Australia, enabling Australia’s research data as a whole to become a nationally strategic resource. It will give an advantage to Australian researchers in three significant ways:

- Research data will be routinely published, enhancing the visibility, reproducibility and reputation of Australian researchers.
- It will be easier for international researchers to work with Australian researchers because of the excellence and visibility of the Australian research data environment.
- New research will be carried out using existing data more effectively and often, exploiting more completely the value of Australia's research data.

The ARDC will be a populated information infrastructure that will achieve the critical mass necessary to become the primary means by which researchers routinely discover and access Australia’s research data. In order to achieve the goals of the Project, ANDS will continue to work closely with providers of research data storage, so that every researcher will be able to manage, store, publish and share their research data. ANDS will also work closely with other NCRIS and EIF-funded capabilities, especially expanding existing relationships with the Atlas of Living Australia, Terrestrial Ecosystem Research Network, the Integrated Marine Observation System, AuScope, and the Australian Urban Research Infrastructure Network.

ANDS has been operating since January 2009. In that time ANDS has built a consensus on the importance of research data and research data infrastructure. In 2010-11, ANDS created a number of national research data services and engaged with a large number of organisations to start the realisation of the Australian Research Data Commons. In 2011-12, ANDS expanded and deepened that institutional engagement, as well as added to the range of available services. In 2012-13 ANDS consolidated its national services, introducing data citation services; continued to partner to build infrastructure through funded projects and selected engagements; strengthened the communities of practice and was a significant leader in policy development nationally and internationally.

In 2013-14, ANDS will build on this infrastructure that has been created. By July 2014, ANDS will have in place additional and enhanced national services:

- a vocabulary service
- citation identifier service
- integrated ANDS systems with ORCID researcher identification service
- Research Data Australia used as a resource by other eResearch infrastructure partners
- functionality added to Research Data Australia to support national collections
- 8 new training programs on ANDS national services
- documentation and training material to support 9 areas of importance to research data management
- improved linkages to publications, services and grants and activities to further improve the context in which data can be discovered.

By July 2014 there will be further support for institutional research data infrastructure in place:
- 20 programs of work to extend infrastructure to support institutional research data assets
- support for showcasing of institutional data assets through the ability to highlight collection strengths
- coordinated 5 community events to support these outcomes.

By July 2014 there will be an established environment of national collections that will have in place:
- 12 collections of national significance from individual institutions
- 4 distributed national collections on themes of research importance in Australia
- established partnerships with RDSI, the nodes and service providers to coordinate access to data and associated services and tools regardless of location
- 9 topic pages in RDA covering research areas of demand.

ANDS will continue to take a leadership role in international data activities. It is a founding partner in Research Data Alliance and will be heavily participating in the second plenary and hosting the third in the coming year. It will also be providing support for the activities of the working groups associated with Research Data Alliance. ANDS will continue its active engagement in the DataCite consortium and further its relationship with ODIN. Additionally ANDS will be more closely coordinating its efforts with its international counterparts specifically in the UK and Netherlands.

These outcomes, in concert with previous year’s activities, will produce the following desirable consequences.

**Better Data**

- A national resource of managed, connected, findable and re-usable research data now exists that previously did not.
- More data are being automatically captured from a wide range of instruments, ranging from small sensors to major national facilities.
- Data providers are increasingly seeing the value in providing their data through richer shared environments.
- Data are now visible to researchers that were never visible before.
- Researchers can more easily find and gain access to public sector data.
Better Research

- A cohort of leading Australian researchers has demonstrated the value of a richer data environment.
- Australian researchers can publish and cite their research data outputs.
- Larger questions and grander challenges can be addressed by a data commons that spans facilities, institutions, disciplines and sectors.
- New types of research are now possible that previously were not.
- Richer data environments can facilitate evidence-based policy.

Better Institutional Capacity

- Data can be captured, managed, and shared more easily at most research organisations in Australia.
- Research organisations and public sector agencies now provide policy support and invest more in data management.
- Australian research institutions can more easily comply with the requirements of the Code for the Responsible Conduct of Research.
- There is an established community of data management professionals.
- Australian researchers and research organisations have better capability to operate and exploit the new research data infrastructure.

Better Investment

- There is more cost-effective reuse of research data software in Australia.
- Australia has an improved ability to measure the quality and impact of all of its research outputs.
- Research Data Australia reduces duplication of effort and cost of unnecessary data acquisition by facilitating re-use.
- Common approaches, standards, and services have increased efficiency and coherence across the national data infrastructure investment.
- Established data management infrastructure now allows for stronger national policy and easier institutional compliance.

Better International Engagement

- Australia is seen to be a leader in research data infrastructure internationally.
Australian researchers can more easily engage with international partners on data intensive research.

As a consequence of these individual outcomes there is an over-arching outcome; Australian researchers now have access to infrastructure that enable them to:

- systematically, reliably and authoritatively connect their research data to project, institutional and disciplinary descriptions
- simultaneously publish citable research data collections through institutional, disciplinary and national services.

This will ensure that Australia has a mature, globally leading capability in research data, making it a key locus for data intensive research. This capability will be demonstrated by leading researchers in a variety of disciplines to show the power of this infrastructure to enhance their research.

### 3 Status of Project

The status of the Australian National Data Service can best be understood in terms of its four sequential stages: establishment, initial NCRIS funding, Super Science funding, additional funding, and time extension.

#### 3.1 Establishment

The Australian National Data Service can trace its beginnings back to the Platforms for Collaboration (PfC) capability as part of the development of the National Collaborative Research Infrastructure Strategy. During the course of the PfC facilitation process, a number of workshops were held to determine the activities that might be included in the investment plan to assist research data management. Following the approval of the overall PfC investment plan by NCRIS, an implementation workshop with wide representation was held to confirm the proposal to establish the Australian National Data Service (ANDS). This workshop took place on May 29, 2007. It endorsed the ANDS concept and proposed that a technical working group (the ANDS TWG) should be formed to draft a more detailed statement on the purpose and goals for ANDS, moving beyond the conceptual definition provided in the PfC investment plan. This working group met both physically and virtually over the course of 2007, and in October produced Towards the Australian Data Commons: A proposal for an Australian National Data Service\(^2\).

In late 2007, the then Department of Education, Science and Training (DEST) asked Monash as the lead agency to work with ANU and CSIRO on a project to take the next step and establish the Australian National Data Service (ANDS). The ANDS establishment project concluded in December 2008.

---

3.2 NCRIS funding

The ANDS Draft Interim Business Plan was submitted in September 2008, and the Interim Business Plan was submitted in December 2008. ANDS commenced officially in January 1 2009. By March 1, 2009, ANDS had 16 staff. NeAT Round 1 projects were underway and a second round of NeAT projects was identified.

3.3 Super Science Funding

In the May 2009 budget, the Commonwealth government announced a series of initiatives collectively labelled as Super Science. The ANDS 2009-10 business plan was submitted in March 2009 (prior to this announcement) and accepted in July 2009 (post this announcement). The substance and execution of this plan was substantially affected by the ARDC project (announced under the Super Science program and funded from EIF). Consequently considerable effort was expended on creating a project plan for the ARDC that was complementary to the NCRIS-funded activities.

The ANDS Steering Committee decided in mid 2009 to recommend to the then Department of Innovation, Industry, Science and Research (DIISR) that ANDS manage the NCRIS-funded and EIF-funded activities as an integrated project. The Steering Committee also decided to reshape the portfolio of ANDS programs to better reflect the implications of, and constraints on, the added funding. As a consequence, the existing separate Frameworks and Capabilities programs were merged, and the Utilities program was moved from NCRIS-funded to EIF-funded and renamed ARDC Core. Four new EIF-funded programs were instated: Data Capture, Metadata Stores, Public Sector Data, and Applications. In the period July 2009-March 2010, ANDS consulted widely on these changed plans, and after some fine-tuning to respond to consultation feedback commenced executing against them.

One of the requirements of the ARDC project was that ANDS would provide to DIISR an initial specification for the ARDC that would detail $10M of early expenditure to support the creation of the ARDC. This required commitment of funds by September 30th 2009. This initial commitment was described in the Preliminary Specifications Report as “early activity” expenditure, but has come to be known as “fast start”. Consequently ANDS produced a description of proposed (now actual) engagements based on discussions already underway and relationships that had already been established. A number of the activities described under various programs in the Business Plan below were funded as early activity/fast start, with two goals. The first was to start expending the allocated funds (which at the time had to be expended by the end of 2010-11), thus smoothing somewhat the expenditure curve. The second was to quickly undertake a range of activities from which ANDS could learn and thereby fine-tune the process of expending the remainder of the Super Science funding.

3.4 Additional Funding

As previously described, ANDS has also entered into three other Funding Agreements worth a total of $462K with the DIISCCRTE. The first agreement was to progress outcomes arising from the First
Joint Australia-EU Research Infrastructure (RI) Workshop in June 2011. Subsequently, in 2012, two more agreements were entered into, for the Second Joint RI Workshop, and the DataWeb Forum. The DataWeb Forum has provided partial funding for Australia’s participation in the establishment and development of the Research Data Forum (previously called the DataWeb Forum).

The most recent funding agreement was to maintain this research infrastructure under the Collaborative Research Infrastructure Scheme (CRIS). This business plan provides a combined view of how ANDS intends to execute all of these activities in 2013-14

### 3.5 Time Extension and Reporting

In April 2010, an opportunity arose to ask DIISR whether a short extension might be possible for ANDS, and they advised that there was an opportunity to extend for a further period of two years to harmonize with other NCRIS and EIF investments. ANDS staff and the Steering Committee managed to identify shifts of funding and timing across reasonably permeable boundaries that still delivered a viable ANDS, one able to continue to deliver on behalf of the Australian research community through to June 2013. This extension of time required a re-allocation of funds between programs, as well as a change to the funding profile within programs. It also required the funding of the project office over a much longer period. Consequently an additional $0.5M was provided by DIISR under NCRIS funding to support the operation of ANDS over a longer period. A three-year high-level project plan was developed, so that ANDS:

- honoured all existing commitments
- continued with existing partnerships - this means continuing to actively engage with the research institutions
- retained the capacity to work with data champions
- maintained an ongoing capability of engaging with the sector.

A further extension was sought in order to most effectively use the CRIS funds, as they were sufficient for only 7 of the 18 month period from July 2013 to December 2014. The project plan shows an uneven 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. This business plan thus describes the proposed activity over the final year of a four year plan concluding in June 2014, and the chart in the finance section shows the intended expenditure pattern for the various programs. This chart shows the expenditure against the newly constituted programs, mapping the previous structure to the new structure.

Another important variation to the ARDC project contract, agreed in March 2011, replaced quarterly milestone reports being delivered individually with annual reporting that incorporates reports on ARDC progress as well as NCRIS progress.
3.6 The Function of ANDS

ANDS exists to transform Australia’s research data environment by making Australian research data collections more valuable though improved management, more richly connecting data, enabling discovery and supporting the multiple use of this data. The purpose of this activity is to enable richer research, more accountable research, more efficient use of research data, and improved provision of data to support policy development.

The ANDS project now runs from 2009 to 2014. The function of ANDS to help transform Australia’s research data environment will not be completed by December 2014. As a consequence ANDS will develop a map of the function of ANDS over 10 years informed by *Towards the Australian Data Commons*. This will determine future directions necessary to ensure the effective delivery of the broader 2011 Strategic Roadmap for Australian Research Infrastructure.

3.7 ANDS Programs and their Transformation

Originally there were nine programs established to conduct the work of ANDS and these have been modified, concluded and reshaped over time into the current four programs:

- **National Collections** of government, institutional, discipline and national facility data, taking over from the Public Sector Data and National Collections programs
- **National Services** including registration, publication, discovery, and advisory services, taking over from the Frameworks and Capability, and ARDC Core programs
- **Institutional Engagement** with all of Australia’s major research organisations to ensure effective research data technologies and infrastructure, taking over from the Seeding the Commons, Data Capture and Metadata Stores programs
- **International Collaboration** with data infrastructure providers to ensure that Australian research data infrastructure is compatible with international approaches.

The ARDC Applications program is concluding, though there is a small amount of remaining activity being managed in the Institutional Engagement Program.

The following table shows the intended size and focus of the programs over the whole funding period. It takes into account interest earned as well as project funding (excluding CRIS funding).

<table>
<thead>
<tr>
<th>Programs</th>
<th>EIF ($M)</th>
<th>NCRIS ($M)</th>
<th>Other ($M)</th>
<th>%</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Collections</td>
<td>6.1</td>
<td>1.3</td>
<td>-</td>
<td>9.8%</td>
<td>ANDS</td>
</tr>
<tr>
<td>National Services</td>
<td>10.3</td>
<td>3.1</td>
<td>-</td>
<td>17.7%</td>
<td>ANDS</td>
</tr>
</tbody>
</table>
### Notes:

- The Total Allocated Funds is $75.5 million which consists of EIF & NCRIS Allocated Funds of $75.3 million (EIF $52.7 million and NCRIS $22.6 million respectively) and Other Funding from 1st and 2nd Joint RI Workshop and Data Web Forum of $0.5 million.
- Total International Infrastructure program planned expenditure for the financial year 2013-14 is $486K, of which $284K is funded via EIF funds. The remaining $202K for this year is funded via 2nd Joint RI Funding ($400K and $45K respectively) and Data Web Forum funding which will also fund activity in 7-12/2014.
- National Collections distributions are through a combination of CSIRO and Monash with Monash disbursing contract payments directly.
- National Services distributions are through a combination of ANU and Monash with Monash disbursing some contract payments directly.

### 3.8 ANDS Principles

In responding to the new objectives and program requirements, ANDS continues to follow its foundational principles:

**Commons Framework:** ANDS has started in a way that anticipates the need to scale up and adapt over time via an extensible framework of data stores, federations and services that enable better data creation, capture, management and sharing.

**Focus:** ANDS will continue to identify and work with those who are ready, willing and able to contribute significantly to the ARDC vision, and who provide the most strategic return to the ARDC for the effort expended. However, ANDS will endeavour to support all of the larger research institutions directly, in order to rapidly achieve critical mass.

**Content:** ANDS is initially focussing on content recruitment into stores and federation across stores so as to achieve a wide coverage of data quickly at an agreed level of quality; in later years the emphasis will shift towards quality improvement.

**Service Provision:** ANDS is focussed on service provision and infrastructure development, not research and exploration; its programs will develop, integrate, and continually improve production-level systems in support of well-understood services.

**Strategic Partners:** ANDS recognises the need to be open to, and engage appropriately with, innovations and external institutions relevant to the ARDC, including the Australian Access
Federation (AAF), the National Computational Facility (NCI), the National eResearch Collaboration Tools and Resources (NeCTAR) and the Research Data Stores Initiative (RDSI).

**Stores:** ANDS assumes an environment where storage and long-term curation occur in nationally or institutionally-supported stores, either existing or brought into being over the life of ANDS. These stores will preferably hold objects described by various discipline-specific and documented metadata schemas. ANDS will work with whatever repositories exist, national, institutional or disciplinary.

**Sustainability:** Research data management requires a long-term commitment. ANDS has developed its plans on the assumption that the current funding does not represent a one-off investment in data. The enduring changes forecast in this document within each program are also intended to be sustainable beyond the end of the ANDS planning period.

### 3.9 Scope

**Constituency:** ANDS works with a variety of publicly funded institutions that produce, manage or consume research inputs and outputs to achieve its aims. The scope includes:

- all Higher Education Providers in Australia
- all research organisations that are publicly funded, including CSIRO, GeoScience Australia (GA), Bureau of Meteorology, Australian Bureau of Statistics (ABS), Australian Institute of Marine Science (AIMS), the Australian Antarctic Division, Departments of Primary Industry
- members of the cultural collections sector (galleries, libraries, archives and museums).

As a component of Platforms for Collaboration (PfC), ANDS is funded to work with all research disciplines in Australia, not just the NCRIS capabilities. This means that the specific concerns of the Humanities and Social Sciences are also taken into account.

**ANDS Community:** The ANDS Community consists of providers of research data and ANDS services, consumers of research data and those services, and managers of research inputs/outputs. This includes key stakeholder aggregations such as CAUDIT (The Council of Australian University Directors of Information Technology) and CAUL (The Committee of Australian University Librarians). The ANDS Community includes the general public only to the extent that they will be able to use some ANDS services to discover and access publicly available data.

**Data:** ANDS is concerned with the digital data that is produced by researchers as well as data that is used by and made accessible to them. Data is the information that researchers study, that is transformed by researchers and produced by researchers. Research publications are not included within the scope of ANDS but files, images, tables, databases, models, computer outputs, and similar digital representations are included. ANDS will support the ability to create links between data, publications, software code and visualisations, where these may appear as either research inputs or research outputs.
3.10 ANDS’ Impact

3.10.1 Research Data Collecting Institutions Overview

Collections Approach
As the concept of data publication and sharing gains traction, and the volume of data available in the Australian Research Data Commons increases, the value of taking a collections approach is becoming apparent. This approach has long been used in libraries to introduce focus and applicability into the presentation of material. Collectors can track content across time series, locations or related subject. The heart of this approach is to partner with institutions to manage their data assets.

In ANDS this approach will allow the highlighting of data collections with an increasing focus on those that are of national significance. It will work with collecting institutions to bring this material together, describe it in a way that enables flexibility in presentation, and highlight it for discovery. Enabling this approach to the management of data provides the following benefits:

- enables increased discoverability and browsability of data collections
- provides a more complete picture of data on any given topic
- shines a spotlight on important and significant data
- enables new and more complex research questions to be addressed.

By working with collecting institutions, ANDS can leverage their understanding of the data and research environment in which it is created and used to develop the data descriptions that enable it to be presented in meaningful contexts.

Collection Types
There are a number of ways in which this collecting may occur.

1. Single Significant National Collections
   These collections are collected and maintained by a single institution or unit. They can be definitive, comprise specimens or observations and may be seminal in nature. Their function may be likened to reference material in libraries and be of broad uptake outside a specific discipline. However there may also be collections in this category that are largely confined to a specific discipline, but may have unexpected uses. Curation of these collections lies with a single institution. They can be seen as a significant institutional asset.

2. Collections brought together for a specific collaboration
   These are often gathered from a range of sources and they could also cross a range of disciplines. They are brought together to address a specific research activity and may be transitory in nature depending on the size and longevity of the research activity. The question of the continuing maintenance of the collection once the collaboration is complete will arise. Datasets may be
repurposed. Curation of these data collections will be as varied as the number of collaborators unless data curation is identified as a specific activity within the collaboration. This may then become an issue when the collaboration is complete and funding for curation ceases.

3. Distributed Collections

One of the values of a national metadata store is that it provides the ability to pull together ‘like’ data. There are two ways this can come about. One is a coordinated approach to collecting from institutions (collecting with intent) in which relevant institutions describe and publish collections with a view to their use. Examples would include the collections that are in discipline portals. Another is to do this on the fly within Research Data Australia using the metadata. This method allows a ‘slice’ to be taken that crosses disciplines and formats and profiles a collection based around a topic or location. Thus a collection could be formed of the data within RDA on water that includes observations, models, photographs, historical anecdotes, commentary on water use, socio-economic impact statements etc. It has the potential to provide wider and deeper insight into a given issue. It also delivers value in the serendipitous exploration of data for innovative research and strongly supports the addressing of the ‘big’ questions: complex research issues. Additionally it presents data in a way that supports e-research methodologies. A characteristic of this type of collection is its fluidity. A dataset can belong to more than one collection and be presented in many contexts.

4. Institutional profiles

While not necessarily a national collection, a collections approach has the value of being able to showcase the data produced by an institution, either as research output or as support for Australia’s research effort. The value of Australia being able to showcase the data outputs of its institutions could contribute to attracting research and funding internationally.

The ANDS National Collections program proposes a series of activities to identify such collections and work with institutions to describe and publish them as collections of national significance. The program recognises the specific subject and content expertise contained in the institutions as well as the coordination capability. It also proposes working with the current ANDS services to allow the profiling of data collections within the Australian Research Data Commons.

5. Locally significant collections

Often collections arise from a single research project or from a single strand of research. Treating the data as a collection enables effective management, connection, discovery and reuse. This generally needs institutional support and has been the basis of engagement with institutions through many of the Seeding the Commons projects and Data Capture projects.

How collections are used is also a factor in defining collections types. The NSF (National Science Foundation) uses the following typology for data collections:

- **Reference data collections** are intended to serve large segments of the research community.
- **Community data collections** serve a single discipline community.
- **Research data collections** are the products of one or more focused research projects and typically contain data that are subject to limited processing or curation.

National Collections would typically fall into the first two categories.

**Collections Engagement**

Through previous ANDS Seeding the Commons, Data Capture, Public Sector Data, and our new National Collections program we have engaged with institutions to support their data collections and to support the role of data collecting. Our focus in the National Collections program will be on those collections of national significance, particularly those collections that may be made available through an RDSI node. This will be complemented by our Institutional Engagement Program focusing on institutional research data assets. In all cases the engagement has been structured to enhance the value of those collections through improved management, connectedness, discoverability, and reuse.

**3.10.2 Research Institutional Engagement Overview - Coherence**

As many institutions are now reaching significant milestones in their data management projects two particular aspects of ANDS’ support for institutional research data management are emerging:

- the continuing relevance of ANDS’ engagement with the institution as a whole
- the importance of institutions undertaking an enterprise approach in a range of areas in order to deliver the ANDS “four transformations” effectively.

At the same time, ANDS is ramping up the provision of services to allow for consistent ways of publishing, discovering, accessing and using data.

In order to bring together these services and the institutional approach ANDS is encouraging a coherent approach to research data infrastructure; without this, Australian research will be unable to reach its full potential.

ANDS plans to engage with its research institutional partners through a coordinated approach that will assist partners to achieve this coherence and realise the following benefits:

- develop knowledge of the full extent of the institution’s data holdings and the subsequent ability to present it as their body of work
- secure valuable data assets for validation and reuse
- improve the institution’s ability to collaborate through a consistent and supported data management infrastructure
- position the institution to be fully acknowledged in the re-use of its data by others and the consequent reputational benefits
- improve the institution’s ability to demonstrate research excellence.
The development of coherent institutional research data management across the sector ultimately supports the building of a sound ARDC infrastructure.

This approach is a way of delivering a combination of services, software, policy and resourcing that will enable a research producing institution to develop and provide the necessary infrastructure so that its researchers can effectively manage their data to be published, discovered, accessed and used.

The aim of the approach should be to present a unified, consistent picture of the infrastructure, policy and procedures that ANDS thinks an institution needs to have in place for effective research data management, publishing and research support. The approach should actively support the four transformations as seamlessly and effectively as possible.

An institution with a coherent approach is one whose data infrastructure can be accessed and used across the organisation, as well as being consistent with infrastructure approaches across the sector. ANDS will continue to work to put in place services that enable and support this form of coherence across the sector. Solutions should not limit the ability of an institution to interact with the broader environment in the sector. The aim of the approach should be to enable every data collection that can be made available to be made available in a timely fashion.

The way the approach is implemented, and the focus areas that the institution chooses to address, should be in line with the larger data aspirations of that institution. That is, they should enable the institution to participate in collections of particular interest or value to them.

ANDS’ role will be in encouraging, supporting and/or funding institutions to have the following coherent capabilities in place:

1. An institution wide data management policy that includes relevant planning options for researchers where appropriate.

2. IT infrastructure for the institution that enables this policy to be implemented. At the most basic level, if there is nowhere to store data, and/or no easy way to collect it into that storage, then there is no point in having a management policy for it. Equally this infrastructure should be capable of connection to IT tools that will allow for sharing, re-use and reconfiguring. These tools may have been developed locally, through ANDS funded programs such as Data Capture or Applications, through other DIISCCRTE funded projects such as NeCTAR, or outside Australia.

3. Co-ordination of the policy and IT infrastructure to encourage and facilitate the creation and storage of data in a fashion that will enable sharing and re-use at an appropriate future time.

4. Storage and collection mechanisms that support the capture and creation of metadata about the data, so that the data can be shared and re-used. Where possible this should be drawn from existing sources (to avoid re-keying) and be as widely shared as possible. From an ANDS perspective, these metadata should be made available in the Australian Research Data Commons.
5. **Dedicated services** in place to support all of the above that draw from the relevant areas of expertise across the institution.

Institutions will implement research data support in a wide variety of ways that will best suit their needs. By seeking common outcomes from similar initiatives at institutions across the sector in Australia, ANDS is seeking to enable the Australian Research Data Commons to be a nationally strategic resource. ANDS will work with institutions in this way through all of its programs but with the engagement being focused through its Institutional Engagement program.

### 3.10.3 National Research Data Infrastructure Partner Overview

ANDS is part of the Government’s investment in research data infrastructure in Australia. This investment can be seen very broadly as comprising all of the investments in research data generating activity – indeed many research grants have a substantial data acquisition component. However ANDS has a particular responsibility and opportunity to engage in partnership with other major national data infrastructure investments. ANDS has to date particularly engaged with:

- some of the major data generating instrument investments including the Australian Synchrotron, the Australian Nuclear Science and Technology Organisation’s neutron beam instruments, the Australian Telescope National Facility, as well as investments in a larger number of smaller data generating instruments through investments like BioPlatforms Australia, the Australian Microscopy and Microanalysis Research Facility, and the National Imaging Facility
- the major problem or discipline focused data investments, including the Terrestrial Ecosystem Research Network (TERN), the Integrated Marine Observation System (IMOS), the Atlas of Living Australia, AuScope, the Australian Urban Research Information Network (AURIN), the Australian Data Archive, the Australian Biosecurity Information Network, and the Population Health Research Network (PHRN)
- the major eResearch enabling data investments including NeCTAR, with investments in data tools and data collaboration, RDSI for data storage, and the high end computation facilities to enable data analysis and generation.

ANDS’ engagements to date and planned engagements have occurred across all of the ANDS programs. ANDS continues to engage with a number of capabilities on data licencing issues through our work with AusGOAL and an ANDS sponsored working group. A number of institutionally based projects in our Seeding the Commons and Data Capture projects have supported the richer capture and management of research data from NCRIS or Super Science funded instruments, for example flux tower data entered using a Monash University metadata access tool into a Metadata Store as part of the TERN OzFlux program. Collections from many of the facilities are discoverable through Research Data Australia and collections can be persistently identified. TERN has launched its own portal based on the ANDS registry tool to expose its own collections. Public Sector Data is funding work to make Geological Data more available to exploit using the AuScope portal and more discoverable using Research Data Australia. ANDS funded the automation of real-time publication of data from Research
Vessel *Southern Surveyor* and Research Vessel *Aurora Australis* available through the IMOS portal and with collections descriptions discoverable through Research Data Australia.

ANDS has partnered with a number of the National Data Investments to fund demonstrations of the value of new approaches to research data through its Applications program. ANDS is partnering with the Terrestrial Eco-Research Network, the Atlas of Living Australia, BioPlatforms Australia, the Integrated Marine Observation System, the National Imaging Facility, the Population Health Research Network and the Australian Urban Infrastructure Research Information Network to together demonstrate these advantages, in particular demonstrating the value of integrating data from different disciplines. Examples include the bringing together of satellite, genetic and field observations of Australian soil, or the ability to visualise and combine biological data though many dimensions from whole-of-organism through microscope level to genetic level. A particularly powerful example is the demonstration of the value of examining very wide scale effects of climate change, response and adaptation though built environment, natural environment and health though sharing climate data downscaled to local conditions using National Computational Infrastructure computation, though a portal at Griffith University, utilising the network of researchers around Australia organised through the National Climate Change Adaptation Research Facility and with the intent to store the data on an RDSI funded node.

Finally ANDS is focusing on nationally significant research collections though partnering with a range of collection institutions with a view to enabling Australian researchers to get better access to this data, often stored and made available through RDSI funding.

ANDS thus has engaged with a very wide range of activities in collaboration with other research data capabilities, but also participates in efforts to ensure that ANDS strong institutional focus complements other more specific investments.

The net effect of all of this activity is that the many data investments that Australia makes are enhanced by having data collectors, research institutions, and research infrastructure providers combine to give Australian researchers a significant research data advantage.

4 Research Infrastructure

For researchers to work in the world of data-intensive research, they will need:

- policies that support a new way of working
- a technical data fabric that enables storing and moving data
- a metadata infrastructure to manage rich information about their data
- a referencing mechanism that enables input data, modelling outputs (such as visualisations), software code and documents to be cross referenced
- the ability to search across all the collections that have been registered
- training and training materials that enable the infrastructure to be used well.
To deliver that infrastructure there is a need for substantial data collections, and a combination of national services and coherent institutional services. ANDS is creating many of those services, and helping to seed institutional services that are optimised to be part of the Australian Research Data Commons where re-use, sharing and commonality of approach is possible because of the coherence achievable with national investments. These services can also be integrated with other national services – whether they be data storage services, high intensity data analysis, or discipline or problem specific data services.

ANDS has instituted four programs to establish and maintain this infrastructure:

- National Collections
- National Services
- Institutional Engagement
- International Collaboration

These programs will be delivered in a co-ordinated manner that will ensure that ANDS’ partners engage with ANDS as a whole, not with the individual programs. This has the consequence that ANDS needs to have a greater emphasis on customer relationship management so that partners do not have to navigate their way through different parts of ANDS.

### 4.1 National Collections

#### 4.1.1 Program Aims

The value of establishing national collections of research data is considerable and addresses many nationally significant data needs. One is to have collections of data that enable big problems to be addressed, often by investigating “survey” questions. Another is to have a reference collection that enables researchers to compare their results with a standard. A further need is to have a locus of all research data of a particular form to be explored. By bringing together ‘like’ data and highlighting collections of national significance, ANDS supports research through the improvement of its discoverability and subsequent potential for reuse. This consolidation also has the potential for showcasing valuable holdings globally and increasing the potential attractiveness of Australian research environment.

There are many different players that have an interest in addressing part of the problem. Through the National Collections program ANDS intends to harness this and increase the focus on establishing a rich set of such collections, leveraging off its many institutional partnerships and proceeding in a manner complementary to RDSI and RDSI nodes activity in storing national collections. The National Collections program will also work with other ANDS programs to effect the proposed changes.

#### 4.1.2 Program Overview

Two ANDS programs, Public Sector Data and National Collections, are combining to form the National Collections program. To date Public Sector Data has engaged with over 19 agencies raise the profile
of data management and expose their collections. It has resulted in over 34,000 collections and datasets being published through Research Data Australia. National Collections has promulgated awareness of the value of a collections approach to data and gained some traction in acceptance of distributed national collections. A proof of concept has been developed and demonstrated with positive results. The merging of these programs is indicative of the importance public data collections play as an input to research and the increasing value a collections approach is going to deliver particularly when combined with the delivery of storage solutions and services in partnership with other organisations.

In leveraging off work that has been conducted through the other programs, notably Data Capture, Seeding the Commons and Public Sector Data, the National Collections program will increase focus on collections of national significance that have already been captured and published via Research Data Australia. It is also working through the engagement process within the Institutional Engagement program to identify and expose further collections for reuse with the research institutions and agencies. Its purpose will be to support the research sector and the research data providers in this identification of nationally significant collections of research data and ensuring the availability of these collections; that they are well managed, connected, discoverable and able to be used as widely as possible for as many purposes as possible.

This program will work with all parts of the research and public sector in determining these national collections. It will work closely with research organizations to identify holdings of national collections. The decision on this significance will lie with the institution. It will also work in partnership with the ReDS program of RDSI together with the RDSI Nodes on approaches to making national collections available and connecting nodes that store national collections to collections descriptions that are made discoverable through Research Data Australia (RDA) and other portals. It will also work with research communities and research data providers on establishing sustainable arrangements for national collections.

Each of these discussions would involve the collection holders and RDSI nodes who might host these collections. The program will provide more focus to this type of work, and provide a greater focus on national collections within the broader Australian data holdings. This greater focus has a number of benefits:

- provide researchers with unprecedented access to national collections that enable new forms of investigation to take place
- provide ANDS’ institutional research partners with a coherent effort that will enable them to build well managed, connected, discoverable and widely usable collections
- provide a focus for discussions with research infrastructure capabilities and research groups on how national collections might best be established and hosted
- help provide a very rich set of national collections that might be hosted on RDSI infrastructure, reducing storage costs and improving accessibility of collections.
4.1.3 Infrastructure Created to Date

The National Collections team has established a framework for the publication of collections of national significance. This includes the identification of the desirable characteristics of such a collection as well as defining processes for the identification and description together with access in a wider context. Two types of collection have been identified for focus: those that are curated and published by a single institution and those that are distributed – the collection is one that brings together ‘like’ data from a number of institutions. Examples of activity from single institutions are the CSIRO National Biological Collections and Geoscience Australia’s ASTER collection. A distributed collections ‘proof of concept’ was developed during the 12/13 period to test the value and explore further requirements for distributed collections. It was done for the Australian Water Research Coalition on the topic of urban water. It has proved to be very successful. One value that has emerged has been to allow data associated with research initiatives to be presented in context and identified and published within the institutions funded to carry out the research. Many of these institutions are already ANDS partners. The Urban Water National Collection will include data published by Bureau of Meteorology, CSIRO, Griffith University, UQ, University of Adelaide, UniSA, Flinders, in the first instance with other universities being approached as part of the institutional engagements as the funding bodies identify their projects. Other activities into new areas will be with state government agencies and utilities. Engagement with identified targets was influenced by their readiness and capacity to work on national collections. As it turned out resourcing within institutions was largely given over to the completion of the ANDS funded projects and the teams in institutions were not ready to focus on further collecting without base infrastructure. The program maintained a relationship with the ReDS program in RDSI, providing feedback on documentation and strategy as required and providing a comprehensive outline of ANDS activities and the potential for intersection and collaboration. The team also met with the nodes to identify high impact areas for collecting. Progress with the establishment of RDSI storage and allocation processes took longer than anticipated which limited ANDS capacity to assist.

The Public Sector Data program has been working with a number of agencies to publish their data. These include Geoscience Australia, Australian Institute of Health and Welfare, Australian Antarctic Division, Bureau of Meteorology, Australian Institute of Marine Science, Murray Darling Basin Authority and ABS. The aim has been to raise awareness of the value to research of publishing data and assist where possible with establishing publication feeds to RDA and relevant discipline portals. The outcome has varied across agencies depending on the complexity and level of description of their data. It has resulted in some high impact publication. The team has also worked with NCRIS partners such as AODN and ALA to establish feeds from their portals. Funded projects out of Public Sector data have resulted in the publication of data in the legal, museum, water resources, oceanographic and geospatial areas of data. Activity in this program has served to establish a foundation for identifying and highlighting collections of national significance from the public sector.
<table>
<thead>
<tr>
<th>Agency or Institution or Project</th>
<th>Project Status and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AuScope SISS Deployment</td>
<td>Deployment of the Spatial Information Services Stack (SISS) into 10 government and state or territory agency data providers including CSIRO, Geoscience Australia, BoM, ODSM and DPI (Vic).</td>
</tr>
<tr>
<td>CSIRO Water Resources Observation Network</td>
<td>5 significant collections of data from the Murray Darling Basin Sustainable Yields projects have been exposed via Research Data Australia and work formed the basis of the data management system in CSIRO.</td>
</tr>
<tr>
<td>AODN Data from National Research Vessels</td>
<td>This project has delivered 900 core data sets captured from research vessels published via the Australian Ocean Data Network portal and Research Data Australia (RDA) with further feeds of a further 14,000 collections from the AODN portal.</td>
</tr>
<tr>
<td>Australian Legal Information Institute (AustLII)</td>
<td>This project delivered 500 data collections which are principally the ‘raw materials’ of most legal research: legislation of all forms; Court and Tribunal decisions; treaties; official materials interpreting legislation; and reports proposing law reform. Exposure to AustLII databases provides the only free access, comprehensive national view of legal data of this nature. Additionally exposure has been given to LawCite, a completely automatically-generated case citation service.</td>
</tr>
<tr>
<td>Museums Metadata Exchange</td>
<td>This project has coordinated metadata from 18 museums across Australia and delivered a metadata exchange which gathered metadata from these museums and automate a feed of 1,000 collections descriptions into RDA.</td>
</tr>
<tr>
<td>Public Records Office of Victoria</td>
<td>A project that is currently in progress to delivered metadata for archives collections initially from PROV and State Records NSW, with Queensland Archives and NAA proposing to follow. The goal is to include all government archives and provide a single point of discovery for Australian Archives for the first time.</td>
</tr>
<tr>
<td>GeoSciences Australia</td>
<td>This engagement has enabled the automated exposure of over 14,000 data collections, including some of national significance which will be the subject of further activity in 2013-14.</td>
</tr>
<tr>
<td>Australian Institute of Health and Welfare</td>
<td>This engagement delivered descriptions for 17 key data collections including the Death Index and labour force statistics.</td>
</tr>
</tbody>
</table>
### Agency or Institution or Project

<table>
<thead>
<tr>
<th>Agency or Institution or Project</th>
<th>Project Status and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian Bureau of Statistics</td>
<td>Extensive engagement with ABS has not resulted in publication of data to RDA. A project with AURIN to utilise the 2011 census data is in negotiation.</td>
</tr>
<tr>
<td>Bureau of Meteorology</td>
<td>The engagement activities have commenced both directly and via the AuScope project with the deployment of SISS. Additionally the Australian Water Resources Assessment data is in draft, prior to publication.</td>
</tr>
<tr>
<td>Australian Antarctic Division</td>
<td>A successful feed has been established publishing 2,000 collections. Data citation is also in the process of implementation.</td>
</tr>
<tr>
<td>Murray Darling Basin Authority</td>
<td>An engagement has commenced with this agency and they are now currently looking at the IT infrastructure to support data management.</td>
</tr>
<tr>
<td>Atlas of Living Australia</td>
<td>A feed of data collections has been established and 490 collections are now published in RDA.</td>
</tr>
<tr>
<td>AIMS</td>
<td>The eAtlas collection is in the process of being published with further engagement to follow.</td>
</tr>
</tbody>
</table>

The two programs working together have also identified some functionality requirements for Research Data Australia that will enable more efficient.

### 4.1.4 2013-14 Activities

The overall goal of the program is to establish new collections of national and institutional significance in partnership with their custodians, ensuring that many of the national collections are available on RDSI nodes with rich data services provided through the NeCTAR cloud.

In order to achieve this, specific actions under this program include the following:

**Engage with selected data custodian on establishing rich national collections**

Working around the initial themes of Urban Water, Climate Change Adaptation, Great Barrier Reef, TERN and ARC/NH&MRC grants, the collections team will engage with research and public sector institutions to identify data collections for publication within the context of the topic and the tools associated with them. It will require the further development of services and processes to support this and cooperative engagement with other programs in ANDS.

**In consultation with research and data providers, develop an inventory of collections both institutionally and nationally significant**

The team will work through the institutional engagement process to identify institutionally nominated collections in research and public sector institutions. It will also work with research initiatives and institutions to identify contributions to nationally significant distributed collections.
themes as outlined. Additionally a review of current holdings in Research Data Australia will be undertaken to identify and highlight any already published data in these categories.

**Liaise with RDSI nodes and data providers to establish appropriate storage and access**

The program will liaise as appropriate to establish a framework for compatibility and interoperability which will include metadata, collection identification and tagging, services and access. It will also leverage existing ANDS services such as DOIs and handles, crosswalks, software (Product/Tools registry) and ANDS data management services for RDSI-initiated storage. The aim of this is to ensure that data can be accessed and used seamlessly from the collections. This will be particularly relevant in the case of distributed collections where the data storage and services may also be distributed.

**For all collections engage with service providers to provide access to tools associated with data collections**

The ability to view the data in its wider context and connect to the relevant tools and services to reuse it is a significant value add. The program will be working with partners to identify, define and connect the collections in areas that include access, licensing and copyright, auto-generated metadata (proprietary vs standards) and workflow. It will also contribute to the development of a process for capture and publication of outputs of use of these services/tools.

**Richly describe national collections and publish in RDA and other relevant portals**

Working through the institutional engagement process the program will assist institutions to describe significance of their collections and display them in context using nested collections and tagging where appropriate. Gold Standard metadata will be encouraged where possible and assistance will be provided to encourage connections to relevant services, tools, publications, other relevant data, activities and parties. Feeds to/from other discipline portals globally will also be encouraged for the same reason.

**Enhance RDA to provide on the fly presentation of tools and services for national collections and click through access to those services**

In order to present national collections, new features will be required in Research Data Australia. The program will be developing specifications which will include end-user tagging to allow contributors to self publish to distributed national collections, optimised faceted searching across metadata fields, links to associated services and workspaces, embedded search that is activated from link then displays results page and improved sharing of records in draft. Additionally the program will address the management of data derived from services and further investigate the use of ‘See Also’ searching to relevant data services.

**Work with NeCTAR and RDSI to ensure that their online services also utilise data discovery in RDA**

As part of the liaison activities the program will work with these ANDS partners to ensure that linkages from their systems access content in RDA for relevant collections, services, activities and parties either via syndication or look-up via their services. The aim is to provide in-context access to data and the associated tools and services making use of all the facilities provided by the infrastructure partners.
4.1.5 Highlights

The Public Sector Data program while engaging with few organisations has had high impact with enterprise feeds delivering a large number of collections. Within this there will be scope for highlighting a number of national significance in the coming period. Being able to showcase important and significant data has the potential for increasing the value of the institution’s data asset.

The development of the Urban Water proof of concept for a distributed collection has been very well received in that research environment. Collecting like data together to provide a richer environment for research to address more complex questions. It has also been identified as a potentially valuable resource supporting a number of research initiatives. That like data may well cross discipline boundaries and allow a greater insight into the issues associated with complex research questions. It may also prompt new questions to be asked. This work sets a blueprint for further activity in other themes.

4.1.6 Challenges

The decision to work with the institutions on the selection of collections may result in some variation over the life of this Business Plan as actual activity will be subject to the preferences of the institution. However ANDS is not in a position to determine research value and the ultimate identification of collections that are truly of national significance will be evolutionary and determined by uptake. In addition, and given that storage for national collections may vary, RDSI collections may be a subset of a wider group of national collections. Research Data Australia should be the point from which a consolidated view of collections of national significance is delivered. ANDS’ approach, which is to engage at the institution level, must effectively complement the RDSI work on national collections.

Sustainability plans for collections identified as national will need to be considered. What obligations would accompany the identification of a ‘National Collection’ are yet to be fully determined and agreed. Early engagements are informing and the obligations are evolving as use cases reveal requirements. Collection providers approach to collecting will vary. This will require some attention to ensure that collections stay current and have appropriate coverage. It can be incorporated into sustainability planning.

Where a collection is distributed, determining agreed curation and custodianship may present challenges. For example, different licensing frameworks could adversely affect how a distributed national collection could be re-used. Again this is an issue that is more effectively addressed in the course of engagements to leverage off actual use cases.

Nationally significant material is in varying stages of digitisation, making broader access difficult. The provision of integrated access to the data and tools across distributed locations will also provide a challenge that will require a cooperative approach to solutions.
4.1.7 End of 2013-14 Outcomes

By the end of the 2013-14 funding period it is anticipated that the establishment of national collections of research data will provide:

- more nationally significant collections with a strong institutional custodian
- a point of access to a wide range of nationally significant data
- links between collections descriptions on RDA and the data held on RDSI Nodes and associated services and tools
- the integration of current collection activity with increased definition to the presentation of collections in RDA
- context around these data collections demonstrating:
  - research activities associated with the collections
  - services and tools available to enable and enhance re-use
  - connections to institutions and disciplines with which they are associated
  - a process for the identification and sustained exposure of such collections.

ANDS will also have examples of national collections that demonstrate the four transformations ANDS espouses:

- having clarity surrounding their management with description, curation and custodianship clearly defined and where applicable complementary storage arrangements with RDSI
- being well connected to provide paths into activities, parties and services that emanate from or are associated with National Collections
- being clearly identified in Research Data Australia with feeds to other appropriate discipline portals to improve their discoverability
- providing paths to improved accessibility via RDSI where appropriate and identifying tools associated with their re-use.

4.1.8 End of 2013-14 Enduring Changes

By the end of the funding period ANDS will have produced the following enduring changes through the National Collections program:

- A national resource of discoverable, connected, accessible, and re-usable research data now exists that previously did not.
- Data are now visible to researchers that were never visible before.
- Data providers are increasingly seeing the value in providing their data through richer shared environments.
- A cohort of leading Australian researchers has demonstrated the value of a richer data environment.
- Larger questions and grander challenges can be addressed by a data commons that spans facilities, institutions, disciplines and sectors.
- New types of research are now possible that previously were not.
- Richer data environments facilitate evidence-based policy.
- Australian researchers and research organisations have better capability to operate and exploit the new research data infrastructure.
- Common approaches, standards, and services have increased efficiency and coherence across the national data infrastructure investment.
- Australia is seen to be a leader in research data infrastructure internationally.
- These collections become a key part of Australia’s contribution to international research.
- There will be new national collections in identified priority areas that enable new questions to be asked addressing nationally significant research challenges.
- Institutions will be increasingly managing their research data as an institutional asset.
- Australia has a leading role in research data infrastructure internationally.
- ANDS will be a trusted partner of research institutions.
4.2 National Services

4.2.1 Program Aims
This program of work aims to provide the technology platform for showcasing collections of national and institutional significance and will ensure that ANDS continues to deliver valued and reliable national services.

Two smaller ANDS Programs have combined to form the National Services Program: ARDC Core, and Capabilities and Frameworks. Together they will contribute to the success of the overall program by:

- supporting new approaches to data-intensive research by strengthening the overall policy context for, and facilitating the emergence of, the ARDC
- improving the level of capability for data management, data-intensive research and associated technologies across Australia by partnering with willing institutions and NCRIS facilities to improve core data [or data management] competencies
- establishing, operating and supporting ANDS IT services.

4.2.2 Program Overview
The scope of this program includes collection, registration, publication, discovery, and advisory services, and high level input to supporting policy. It addresses several systemic obstacles to the emergence of the ARDC: policy and licensing irregularity; human capability constraints; and the necessary technical and 24x7 operational services so that the content in repositories can be findable, re-usable and linkable.

Core National Services are fundamental research data services that support a cohesive network of data collections and enable discovery, access and other value-add services across the resulting data commons. A technical consultancy is available to assist with integrating distributed data commons infrastructure at research and government instrumentality repositories with core ANDS utilities. Within the technical services capacity there are three areas of activity:

1. A range of utility data services at a sector-wide level (such as cross-discipline discovery services, national collections registry, persistent identifier service, etc). As appropriate these utility services either aggregate information nationally or provide component services across several NCRIS domain areas, Super Science facilities, and other research organisations and communities. Robust infrastructure is established together with a service delivery framework that defines the roles and responsibilities of those providing, supporting and participating in the services.

2. ANDS is commissioning the establishment of a distributed set of data commons infrastructure typically run and operated by government agencies and research
organisations. This information infrastructure allows elements of the research data commons to be connected in an information mesh of interconnected references to the people, organisations, projects, fields of research, and locations related to research data.

3. ANDS offers specialist technical advice and consultancy services around establishing data federation utility services. This advice and assistance applies to research organisations, facilities, and government agencies establishing their own data commons infrastructure and others seeking to make use of and integrate with ANDS core infrastructure and software.

The program provides materials and training to support research organisations and research groups to have effective data management practices around the full data lifecycle. The common approach to addressing these generic issues is to partner with collaborators around specific solutions, including eResearch support groups, eResearch infrastructure groups and research leaders and funding agencies, with associated activities working towards harmonising and streamlining the overall data management framework within which a research data commons can operate.

The program engages with ANDS primary collaborators such as:

- Institutional data holders (CSIRO, NCRIS Capabilities, National Library of Australia, National Archives of Australia, Departments of Primary Industry, GeoScience Australia, Australian Bureau of Statistics, the university sector etc.)
- National initiatives such as the National Committee for Data for Science
- Cross-governmental groups such as Australian Government Information Management Office (AGIMO), AusGOAL, Office of Spatial Policy (OSP) and the Australian Spatial Consortium
- Research funding departments such as Department of Innovation 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)
- The Office of the Australian Information Commissioner
- Sector peak bodies such as The Committee of Australian University Librarians (CAUL), The Council of Australian University Directors of Information Technology (CAUDIT) and Universities Australia
- Discipline leaders within institutions
- Research office staff at institutions.

National Services activities will help build institutional capability to contribute to and better exploit national data infrastructure. Various activities will 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 will 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.
National Services activities lead to services such as consultancy, informal knowledge transfer, workshops, documentation, and training materials both directly and by reinforcing local services. Staff from this program work within an integrated engagement activity with staff from all other ANDS programs and also within the community of e-Research support services. These groups themselves are engaged in capability building within their own institutions as well as with their own staff.

4.2.3 Infrastructure Created to Date

In order to improve the policy environment and the capability to enhance and exploit the research data commons, a range of infrastructure initiatives has already been undertaken:

**Table 3: National Service infrastructure and status**

<table>
<thead>
<tr>
<th>National Service</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Data Collections Registration Infrastructure: Systems to</td>
<td>Production</td>
</tr>
<tr>
<td>allow information about research data collections to be registered</td>
<td></td>
</tr>
<tr>
<td>and maintained dynamically through automated harvesting systems</td>
<td></td>
</tr>
<tr>
<td>National Data Collection Discovery Infrastructure: The Research</td>
<td>Production</td>
</tr>
<tr>
<td>Data Australia portal that aims to provide a comprehensive window</td>
<td></td>
</tr>
<tr>
<td>onto the Australian Research Data Commons</td>
<td></td>
</tr>
<tr>
<td>National Data Collection Page Creation Infrastructure: Allowing</td>
<td>Production</td>
</tr>
<tr>
<td>web pages to be created to advertise every registered collection</td>
<td></td>
</tr>
<tr>
<td>Dataset Identifier Infrastructure: Allowing unique, persistent, and</td>
<td>Production</td>
</tr>
<tr>
<td>internet resolvable identifiers to be minted for objects in the</td>
<td></td>
</tr>
<tr>
<td>research data commons</td>
<td></td>
</tr>
<tr>
<td>Place Names Infrastructure: Web-service enabled gazetteer of</td>
<td>Production</td>
</tr>
<tr>
<td>Australian places established in partnership with OSDM/ GA</td>
<td></td>
</tr>
<tr>
<td>Researcher Identification Infrastructure: Infrastructure to enable</td>
<td>Production</td>
</tr>
<tr>
<td>access to definitive source information and identifiers for</td>
<td></td>
</tr>
<tr>
<td>Australian researchers and research organisations (established with</td>
<td></td>
</tr>
<tr>
<td>NLA) providing valuable information context to the ARDC.</td>
<td></td>
</tr>
<tr>
<td>Research Activity Information Infrastructure: Web service enabled</td>
<td>Production</td>
</tr>
<tr>
<td>information system publishing definitive source information and</td>
<td></td>
</tr>
<tr>
<td>identifiers for all funded Australian research projects (established</td>
<td></td>
</tr>
<tr>
<td>with ARC and NHMRC) providing valuable information</td>
<td></td>
</tr>
<tr>
<td>National Service</td>
<td>Status</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>context to the ARDC.</td>
<td></td>
</tr>
<tr>
<td>Standard Vocabularies Infrastructure:</td>
<td>Prototype</td>
</tr>
<tr>
<td>Infrastructure enabling the web-service publication, management and creation of</td>
<td></td>
</tr>
<tr>
<td>standardised vocabularies used in research data laying a foundation for better</td>
<td></td>
</tr>
<tr>
<td>linkages between datasets</td>
<td></td>
</tr>
<tr>
<td>ARDC Infrastructure Establishment:</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Setting up best practice procedures for commissioning, establishment and</td>
<td></td>
</tr>
<tr>
<td>management of national IT infrastructure commissioned through the ARDC project.</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: National Data Policy Services and areas supported

<table>
<thead>
<tr>
<th>National Data Policy Services</th>
<th>Existing Policy Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Commons Re-Use Policy</td>
<td>ARC, NHMRC, DIISCRTE: invited contributor to developing open access policy for publicly-funded research (publications and data) through ARCom and RDIC (and Dept. Innovation). Open Access (research data) briefs (multiple) to ARC, NHMRC, CAUL, CAUDIT, ARMS</td>
</tr>
<tr>
<td></td>
<td>Membership of sector and government policy committees (the Cross Jurisdictional Chief Information Officers’ Committee - CJCIOC, National Committee for Data in Science - NCDS, Commonwealth Spatial Data Management Group - SDMG)</td>
</tr>
<tr>
<td></td>
<td>Invited contributor to several issues papers from OAIC (Understanding the value of public sector information in Australia, Principles on open public sector information)</td>
</tr>
<tr>
<td></td>
<td>Submissions to inquiries and position papers including Review of Health and Medical Research in Australia, AGIMO Big Data Strategy Issues Paper, ALRC Copyright Enquiry, DIISRTE (e.g. Patents and Copyright, APS200 (The Place of Science in Policy Development in the Public Service), RDIC etc),</td>
</tr>
<tr>
<td></td>
<td>Review of Australian Universities IP policies (internal use only)</td>
</tr>
<tr>
<td>Licensing Frameworks – implementation</td>
<td>Active membership of Commonwealth AusGOAL practitioners working group</td>
</tr>
<tr>
<td></td>
<td>Membership of AusGOAL oversight committee</td>
</tr>
<tr>
<td></td>
<td>Establishment of licensing working groups for university and research sectors</td>
</tr>
<tr>
<td>National Data Policy Services</td>
<td>Existing Policy Support</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Creation of licensing area on ANDS website and ANDS guides, supported by a survey-based FAQs on the AusGOAL website</td>
<td></td>
</tr>
<tr>
<td>Content Providers’ Guide and other associated ANDS Guides Training, events</td>
<td></td>
</tr>
<tr>
<td>ANDS Guides and web materials Training, events International Consortium website with DataCite</td>
<td></td>
</tr>
<tr>
<td>ANDS Guides and web materials Training, events</td>
<td></td>
</tr>
<tr>
<td>ANDS Guides and web materials Training, events</td>
<td></td>
</tr>
<tr>
<td>ANDS Infrastructure Usage Support</td>
<td></td>
</tr>
</tbody>
</table>

4.2.4 2013-14 Activities

2013-14 activities will build on the work done to date, continuing the emphasis on data management capability, data citation, and licensing. New foci emerging for 2013-14 include: alliance with sector peak bodies on data sharing policy; an increasingly matrix approach to supporting people and organisations to build the research data commons; and continuing to improve services which facilitate publishing, discovering, linking and re-using research data through a national infrastructure.

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

- Support complex national collections requirements:
  - Annotations for distributed collections
  - Display of duplicate records
- Collection structural hierarchy and connections viewer
- Support crosswalks from community schema to make collection registration easier
- Establish processes for all ARC/NHMRC grant information to be easily linkable to data collections information (UI and API)
- Test processes for journal publications information to be easily linkable to data collections information
- Optimise search engine indexing
- Continue to support and operate the harvesting and registration service
- Continue to operate the discovery portal
- Provide overview of the national data publication system
- Continue to support and operate the DOI service
- Work with publishers, data archives, and citation indexes to establish a data citation as a routine scholarly practice
- Optimise self service and automated processes for DOI updating (persistent identification)
- Test integration of registry and DOI services (simultaneous mint and register)
- Establish national vocabulary catalogue
- Establish national vocabulary service
- Launch, procedures, promotion and support
- Integrate ANDS systems with ORCID
- Support the use of researcher identifiers and good practice encoding
- Test linkages with Google Scholar

**Provide solutions and support for data portal software**
- Component plug and play architecture
- Community support and communications

**Support the uptake of ANDS Services**
- Support partners providing information about their data assets to the national registry and Research Data Australia
- Support partners implementing data publication and promoting data citation through the use of the ANDS DOI Service
- Support the establishment of the ANDS vocabulary service and support for the initial adopters group
- Work with publishers, data archives, and citation indexes to establish data citation as a routine scholarly practice

**Provide a national capability enhancement service for the data lifecycle focusing on data management and data publication**
- Support data management through the provision of documentation, materials, resources, briefings and educational materials
- Stage events (virtual and face to face) to build capabilities of staff and organisations to undertake data management
- Support the development of self-sustaining communities of practice
- Raise awareness of international trends and practice within the Australian data community
- Provide mechanisms for internal ANDS capability building, including
  - documentation, resources, briefings specific training
  - project/engagement support: high level support, collaboration, team based activity

**Provide a national policy service aimed at engaging with all stakeholders and key decision makers so that coherent policy on research data access is achieved through:**

- Provision of submissions:
  - The key strategic submissions are those invited by the Department of Innovation, and include Research Data Infrastructure Committee and ARCom, as well as *ad hoc* briefings to various business units in the department. e.g. National Services has been requested by the Department of Innovation to draft the OA policy to data, which involves (a) development of an open data policy, drivers for the policy, as well as enablers, barriers, previous investments, (b) Mechanisms for implementation, (c) Definitions and (d) case studies.
  - Other submissions include, for example, AGIMO, Australian Law Reform Commission, ANDS Steering Committee, ARMS, CAUL, UA, ARC and NHMRC, the latter two usually on a semi-regular basis, e.g. Advocating the use of a commissioned report on data management from an institutional perspective to help ARC and NHMRC implement their emerging open access agendas.
- Contribution to committees and working groups:
  - The key committees include RDIC and ARCom (due to their high level strategic/national focus);
  - AusGOAL Steering Committee and the Spatial Data Management Committee
  - Membership of sector and government policy committees (the Cross Jurisdictional Chief Information Officers' Committee - CJCIOC, National Committee for Data in Science - NCDS, Commonwealth Spatial Data Management Group - SDMG)
  - Other important working groups include the Research Data Licensing Working Group for Infrastructure Providers, the University Data Licensing Webinars and several focus groups within the Office of the Australian Information Commissioner.
- Support for institutional engagements: Development is underway of a simple and intuitive way to provide ANDS staff with fully searchable access to all of the main documents,
submissions and reports referred to in the last two dot points, as well as those created previously.

Some unallocated budget for 2013-14 activities is retained to enable ad hoc opportunities and activities to collaborate with institutional and national players, (e.g. universities, NCRIS facilities, state-based eResearch organisations) and international players (e.g., Digital Curation Centre UK) over the course of the year. While largely resourced from existing program allocations, some of these joint projects will involve:

- New resources for collaborative activities such as joint materials production, workshops, pilot projects and tools specification. These will not be allocated through open call, but rather through a "ready, willing, and able" filter with the particular stakeholders mentioned above. These are tentatively scheduled for Semester 1, 2014
- Possible responses to any requirements that arise from partner adoption of the ANDS software products or cloud deployments of ANDS services
- The possibility of new requirements related to aggregation and harmonization efforts with respect to an emerging international research data infrastructure

4.2.5 Highlights

In its first four years of operation ANDS has established a national registry of research data, a national discovery portal, and a data publication identifier service. These are nationally-focused, value-adding services which are more efficient to run once for the sector. For example ANDS has joined the global DataCite infrastructure consortium and provides Digital Object Identifiers for data to the whole Australian research sector. Through the EIF and NCRIS infrastructure, many research organisations have established data archives and metadata provider infrastructure to feed the ANDS national registry, which aims to provide a national cross-discipline view of the data outputs from (and inputs to) Australian research.

The NCRIS project has enabled ANDS to establish a capacity to support, advise and exchange information about data infrastructure establishment at Australian universities as well as targeted assistance for those using ANDS national services. ANDS also has established an expertise and advice service around questions of national policy promoting re-use of research data.

ANDS has made significant progress in supporting research organizations and public sector agencies building the Australian Research Data Commons. Specific highlights include:

- ANDS support forums, guides, materials, and events in a number of key areas:
  - Licensing Frameworks -implementation
  - Collection Description Publication
  - Data citation
  - Ethics
Institutional research data management infrastructure development

Data management plans

ANDS Infrastructure Usage Support

AND is also actively engaged with the ARC, NHMRC and DIISRTE and contributes to a number of public sector data policy and implementation groups such as the Spatial Data Management Group, the Australian Governments Open Access and Licensing Framework steering committee, and the Office of the Australian Information Commissioner’s working group on valuing PSI information.

ANDS has established significant national services to underpin the Australian Research Data Commons. Highlights include:

- National Data Collections Registration Infrastructure
- National Data Collection Discovery Infrastructure
- National Data Collection Page Creation Infrastructure
- Dataset Identifier Infrastructure
- Place Names Infrastructure
- Researcher Identification Infrastructure.

In broad terms, the base national infrastructure now exists for researchers and research organisations to publish, discover and re-use research data with richer contextual information. Potential now exists for tracking and acknowledgement of re-use, but further infrastructure establishment and policy development are required into the future.

4.2.6 Challenges

Effective coordination and embedding of this program of work within the activities of the ANDS institutional infrastructure projects is a high priority, as is strategic collaboration with NCRIS facilities. Key challenges for National Services Program activities are how to work effectively with and through collaborators, and for ANDS to be an effective voice for research data.

The success of the ARDC depends on favourable adjustments to the policy settings throughout the data and innovation sector and in all cases these policy decisions are taken by research organisations or government bodies and not by ANDS. The ANDS role in this area is that of advocate and catalyst.

ANDS needs to work towards a truly sustainable national discovery service. This will involve meeting the following challenges:

- Take software custom-made for ANDS needs and transform it to be used by a community of adopters and contributors
- Support development partners and applications within ANDS projects and engagements for the ARDC Core Infrastructure
- Implement transitional arrangements for post 2014 service provision
The creation and empowering of a community of data management practice will engage all Australian research organisations, who are the right entities to address questions of data capability at the individual, collective, and institutional levels. ANDS will set an agenda for capability building, provide resources to support institutions’ aspirations, and create the right environment for sharing amongst a community of practice. ANDS prefers not to act independently, so a major challenge for this program is to act effectively through that community.

4.2.7 End of 2013-14 Outcomes

By the end of 2013-14 it is anticipated that the National Services Program will have produced the following infrastructure outputs, with corresponding support mechanisms to ensure the widespread uptake of those outputs:

- Enhanced the discovery and descriptions of collections within Research Data Australia e.g. enabling the easy linking of data collections with ARC/NHMRC grant information
- Leveraged various technical capabilities of the National Data Collections Registration to allow integration with external systems e.g. Dataset Identifier Infrastructure, simplification of data registration workflows
- Extended the widespread integration and use of existing ANDS software through outward-focused developer engagement and engaging with other national research infrastructure projects (including NeCTAR and RDSI)
- Provision of integrated materials, events and support for research data management issues of strategic importance to institutions including: ethics, sharing, publication, citation, planning, researcher identification, reuse
- Progressed discussions with peak sector bodies and research funding agencies on the principles and guidelines for the dissemination, management, and re-use of research data.

4.2.8 End of 2013-14 Enduring Changes

This program of work will provide the technology platform for showcasing collections of national and institutional significance. It will ensure that ANDS continues to deliver valued and reliable national services. These enduring changes will be supported by sound and consistent policies. By the end of the 2013-14 funding year, ANDS will have contributed toward, or brought about, the following enduring changes through the National Services program:

- Researchers will be able to routinely publish the data outputs from publicly funded projects by using citation and publication services.
- New types of research are now possible that previously were not.
- Data are now visible to researchers that were never visible before.
- Researchers can more easily find and gain access to public sector data.
- Australian researchers and research organisations have better capability to operate and exploit the new research data infrastructure.
- Australian researchers can publish and cite their research data outputs including through the provision of rich data capture tools.
- Institutions will be able to manage their research data as an asset.
- Data can be captured, managed, and shared more easily at most research organisations in Australia.
- Research organisations and public sector agencies now provide policy support and invest more in data management.
- Australian research institutions can more easily comply with the requirements of the Code for the Responsible Conduct of Research.
- There is an established community of data management professionals.
- A community of practice with a sense of common national purpose has emerged amongst key stakeholders of the ARDC.
- Fundamental issues of data management and usage are clearly recognised and addressed in institutional policy and practice.
- A national resource of managed, connected, findable and re-usable research data and infrastructure now exists that previously did not.
- Larger questions and grander challenges can be addressed by a data commons that spans facilities, institutions, disciplines and sectors.
- Richer data environments can facilitate evidence-based policy.
- Australia has improved ability to measure the quality and impact of all of its research outputs.
- Research Data Australia reduces duplication of effort and cost of unnecessary data acquisition by facilitating re-use.
- Larger questions and grander challenges can be addressed by a data commons that spans facilities, institutions, disciplines and sectors.
- There is more cost-effective reuse of research data software in Australia.
- Common approaches, standards, and services have increased efficiency and coherence across the national data infrastructure investment.
- Established data management infrastructure now allows for stronger national policy and easier institutional compliance.
- By contributing to international consortia and collaborations, this program will build and take advantage of a global data publication infrastructure.
- Australia is seen to be a leader in research data infrastructure internationally.
- Australian researchers can more easily engage with international partners.
- Strong international technical partnership enables Australian infrastructure to leverage international research infrastructure networks.

4.3 Institutional Engagement

4.3.1 Program Aims

This program aims to undertake a series of data-focused engagements with all of Australia’s major research organisations. These engagements will assist the organisations to achieve their research data ambitions. This program takes over the activities of the Seeding the Commons, Data Capture, Metadata Stores, and the final activities of the ARDC Applications programs.

4.3.2 Program Overview

This program is building on the activity that took place in previous years under the Seeding the Commons, Data Capture, Metadata Stores and Applications programs. The aim of the Seeding the Commons program was 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. The aim of the Data Capture program was to simplify the process of researchers routinely capturing data and rich metadata as close as possible to the point of creation, and depositing these data and metadata into well-managed stores. Metadata will need to be held at both collection and object level in order to support re-use. The aim of the Metadata Stores program was to assist institutions and disciplines to better manage the collection and object level metadata associated with research data outputs and associated entities.

4.3.3 Infrastructure Created to Date

As a result of these programs, by the start of the planning period (July 2013) all Australian Universities will have undertaken some Seeding the Commons activity (a small number of projects will be completing just after the planning period commences), the most research intensive will have implemented a number of data capture projects, and a majority will have put in place a metadata store for descriptions of research data outputs. By July 2013, ANDS will have funded the development of wide range of infrastructure components in the majority of Australia’s research institutions. All Australian Universities will have undertaken some Seeding the Commons activity (a small number of projects will be completing just after the planning period commences), the most research intensive will have implemented a number of data capture projects, and a majority will have put in place a metadata store for descriptions of research data outputs. The number of projects
ranges across a significant variety of research endeavours and the extent and diversity is illustrated in the outlining the projects below, grouped by research institution.

In addition, ANDS has been continuously working to grow the community of data managers in Australia through the provision of well-used communication tools (message board, email list and website, including a directory of ANDS funded projects). Additionally training programs have been held to promote the understanding of ANDS’ expectations around data management. This work has been done in conjunction with the Capabilities function in the National Services program. ANDS has also been actively seeking to broaden the discussion through a number of ANDS Community Events but also through the use of video conferencing technologies. Informal state based sessions have also been held to help grow the community and promote information sharing.

So, what has been the result of all this activity? Table 6 summarizes all of the projects undertaken by the Seeding the Commons, Data Capture and Metadata Stores programs, grouped by institution.

Table 5: Infrastructure created by institution

<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Name</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian Catholic University</td>
<td>Data Management Plan and Policy</td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td>Australian National University</td>
<td>Earth Sciences</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td></td>
<td>Optical Astronomy (Skymapper)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phenomomics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Humanities and allied disciplines</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The ANU Seeding the Commons Project</td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td></td>
<td>A Scientific Workflow System for Assessing and Projecting the Health Impacts of Extreme Weather Events</td>
<td>The project will develop software that enhances existing research data infrastructure with tools that merge population, health and environmental data for analysis and inference in environmental epidemiology. It will build a system for the kind of data sourcing, manipulation and analysis that is required for greater efficiency in many epidemiological research areas.</td>
</tr>
<tr>
<td></td>
<td>ANU Data Gateway</td>
<td>This project to develop meta data resources within the University is part of a larger strategy of developing our knowledge management capabilities.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>The Australian Nuclear Science and Technology Organisation (ANSTO)</td>
<td>Scientific Information Architecture</td>
<td>ANSTO used the project to kickstart the process of sharing ANSTO's publicly funded research data.</td>
</tr>
<tr>
<td>Bond University</td>
<td>Data Management Plan and Policy</td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td>Central Queensland University</td>
<td>Centre for Environmental Management Core Data Curation project</td>
<td>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.</td>
</tr>
<tr>
<td>Charles Darwin University</td>
<td>Data Management Plan and Policy</td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td>CSIRO</td>
<td>Seeding the Commons: Enabling CSIRO’s Biological Collections for the ARDC</td>
<td>The aim of this project was to identify biological collections managed by CSIRO that are potentially of significance to the broader research community and to enhance the visibility of those collections. From an audit of collections managed by CSIRO, approximately 40 biological collections were identified.</td>
</tr>
<tr>
<td>CSIRO</td>
<td>ATNF Pulsar Data Management Project</td>
<td>The ANDS-CSIRO-ATNF Pulsar Data Management Project enables the discovery of, assessment of and access to Pulsar data observed at the Parkes Telescope.</td>
</tr>
<tr>
<td></td>
<td>Sensor Data management</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| ALA Bird Distribution | In conjunction with ANDS Project AP30 this project will collaborate closely with Atlas of Living Australia (ALA) to develop a tool that reuses data available with ALA and the JCU Tropical Data Hub to:  
explore with the potential impacts of climate change on a wide range of species in Australia; and  
Engage in improving our understanding of the species and the modelling of species distributions.  
The tool will be accessible from the JCU Tropical Data Hub which is an aggregator to serve Tropical research data sets and models from a single virtual location. |                                                                                                                                                                                                                  |
| SEQUITOR: A demonstration integrated coastal knowledge platform project initiation | This project will develop data integration and visualisation infrastructure for a Coastal Knowledge platform for South-east Queensland known as SEQUITOR. SEQUITOR will integrate observational (e.g. monitoring) data from catchment, river and coast with models of catchment, estuarine and coastal processes to enable the discovery and generation of new knowledge and better understand the catchment-to-coast system. |                                                                                                                                                                                                                  |
| Organisation Metadata Central | The project (Organisation Metadata Central –OMC) will produce and deliver software technologies that will enable CSIRO to capture and update standard metadata fields relating to CSIRO’s projects on project commencement/approval.  
The solution will extract information from several authoritative sources such as SAP, allow the user to enter additional information on to standard metadata schemas. |                                                                                                                                                                                                                  |
<p>| Curtin University of Technology | Building software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used. |                                                                                                                                                                                                                  |</p>
<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Name</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deakin University</td>
<td>Filtration Membrane Fouling Data Collection for Water Treatment Research</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata.</td>
</tr>
<tr>
<td></td>
<td>Crystal Orientation Data Collection for Conversion to a General Data Type</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td></td>
<td>Description and discovery of research data collections available at Deakin University</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metadata Stores Support activity</td>
<td>This metadata store project will deliver an integrated metadata store that draws on information held within Deakin’s core business systems and national sources of truth. Options for the delivery and ongoing use and support of a metadata store at Deakin will be assessed to ensure that the best possible solution is implemented that encompasses the strategic vision of the University and supports the Deakin University research systems roadmap. A technical solution will then be implemented along with processes, workflow and researcher training to support the ongoing capture of metadata.</td>
</tr>
<tr>
<td>Edith Cowan University</td>
<td>Data Management Plan and Policy</td>
<td>This developed university wide data management policies and procedures delivered internal training programs and resources for research data management</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------</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>This work has made the AusStage database available to all potential researchers, without requiring expertise in database manipulation and extraction. This increases the audience of AusStage dramatically, providing a reusable and human readable form of the AusStage data corpus. The ANDS SC22B MPPDA (Motion Picture Producers and Distributors of America, Inc.) project has succeeded in making 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.</td>
</tr>
<tr>
<td></td>
<td>Automated measurement of the responses of wildlife populations to climate change</td>
<td>The particular project has enhanced the effectiveness of the research data for appropriate analyses. Specifically it has delivered more efficient cleaning software for data files collected from lizard data loggers. The project has also delivered a solution for storage of the lizard data files. This will allow much more effective data sharing among the current group of researchers, and among others who follow in the project.</td>
</tr>
<tr>
<td>Flinders University</td>
<td>Metadata Store Upgrade</td>
<td>This project will expand on previous work done to establish a ReDBox based Flinders University metadata store. By building linkages between ReDBox and existing University systems, this project will enable a sustainable, researcher focussed approach to the description of research datasets and the subsequent exposure of the metadata via Research Data Australia and other relevant registries. This project will also deliver workflows and templates to help embed good research data management into scholarly practice.</td>
</tr>
<tr>
<td>Griffith University</td>
<td>Identifying and describing Griffith University's research datasets and making metadata available</td>
<td>The aims of this project were: - Assess research data resulting from ARC, NHMRC and Australia Arts Council grants; - Identify and describe the data and consult with researchers to identify appropriate access to their research data Results: The successful import of 1144 registry objects published.</td>
</tr>
<tr>
<td></td>
<td>Smart Water</td>
<td>Water end use study, data capture from instruments, reporting on domestic water use. This project delivered the</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Smart Meter Information Portal</td>
<td>Smart Meter Information Portal which stores and processes all total water consumption data downloaded from 250+ homes that are being metered and logged as part of the South east Queensland Residential End Use Study (SEQREUS).</td>
</tr>
<tr>
<td></td>
<td>Adult Stem Cell &amp; Neurobiological</td>
<td>To develop a software system to centralise the management of a large volume of microscopy image and related experimental metadata, allowing researchers within the National Centre for Adult Stem Cell Research (&quot;NCASCR&quot;) to more effectively organise and analyse their biological imaging experiments.</td>
</tr>
<tr>
<td></td>
<td>Instrumentation and Research Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Management</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Climate Change Adaptation</td>
<td>This project aims to build a software system that acts as a central information hub for researchers in the Climate Change Adaptation research domain.</td>
</tr>
<tr>
<td></td>
<td>Information Hub</td>
<td>The system will be built to allow users to deposit research data with associated metadata descriptions into a central managed storage infrastructure. The system will also enable a variety of search types (including spatial and temporal searching) of metadata in addition to other discovery tools to locate relevant research data stored in the information hub.</td>
</tr>
<tr>
<td></td>
<td>Griffith Research Hub</td>
<td>With the Metadata Stores Project funded by ANDS, Griffith eResearch Services will further implement and develop the Griffith Research Hub to build on functionality already implemented in previous ANDS and Griffith funded Projects. Full scale creation, management and workflow of metadata records will be fully operationalised.</td>
</tr>
<tr>
<td>Griffith University</td>
<td>National Linguistics Corpus</td>
<td>Establishment of an Australian National Corpus that: - Aggregates data from existing corpora residing at Australian universities, allows the discovery, access and deposition and textual annotations to written and spoken data</td>
</tr>
<tr>
<td>Macquarie University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>James Cook University</td>
<td>Tropical Data Hub</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Tropical Data Hub</td>
<td>Tools Development</td>
<td>This project will create a set of complementary tools consisting of server side data manipulations and web accessible interfaces that are designed to process data stored in the TDH and related repositories. The tools will in essence be a synthetic data product to solve specific scientific and socio-economic questions related to land use and conservation in the tropics.</td>
</tr>
<tr>
<td>Bird Species Distribution</td>
<td></td>
<td>This project will collaborate closely with Atlas of Living Australia (ALA) to develop a tool that reuses data available with ALA and the JCU Tropical Data Hub to: explore with the potential impacts of climate change on a wide range of species in Australia; and Engage in improving our understanding of the species and the modelling of species distributions. The tool will be accessible from the JCU Tropical Data Hub which is an aggregator to serve Tropical research data sets and models from a single virtual location.</td>
</tr>
<tr>
<td>Systematic JCU Tropical data collection discovery and description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCU Research Data Catalogue</td>
<td></td>
<td>The aim of this project will be to select and implement a Research Data Catalogue at James Cook University (JCU) as a university-wide Metadata Solution. The solution will be integrated into the Tropical Data Hub architecture as well as investigate internal data flow and data sources in current corporate, research and IT systems to inform efficient and effective systems integration at JCU.</td>
</tr>
<tr>
<td>La Trobe University</td>
<td>Archaeological Database Development: The People and Place Project</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>The Centre for Materials and</td>
<td>The major outcome from the researchers' point of view is that the system provides a facility to store their datasets,</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Surface Science’s</td>
<td>Remote Laboratory Instrumentation</td>
<td>describe them, and easily share them online, with persistent identifiers so that they can be cited conveniently and expose these datasets to the wider world through the Research Data Australia portal.</td>
</tr>
<tr>
<td>La Trobe University</td>
<td>Metadata Store</td>
<td>This project will contribute to La Trobe University’s research data infrastructure by establishing mechanisms to: manage descriptions of data collections held at La Trobe enable discovery and reuse of data collections held at La Trobe support strategic planning for research at La Trobe ensure high quality of metadata, in support of the other requirements for a metadata store.</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Glycomics Repository</td>
<td>This project advanced the state of data capture and management within the glycomics discipline by installing new repositories, create an automated data and meta-data capture system at the Australian Proteome Analysis Facility (APAF), with an OAI-PMH feed will be created to allow Research Data Australia to harvest details of the collections stored within the GlycoSuiteDB repository.</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Papyri Data Capture</td>
<td>Building software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Seeding the Commons</td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td>Macquarie University</td>
<td>Primary Production in Space and Time</td>
<td>The project will fuse disconnected data sources—weather, remotely sensed land-surface observations, CO₂ and water flux measurements, hydrograph data and remotely sensed CO₂ concentrations—to generate a time-varying field of gross primary production (GPP, the most fundamental function of all ecosystems) across the Australian continent.</td>
</tr>
</tbody>
</table>

Primary Production in Space and Time

The project will fuse disconnected data sources—weather, remotely sensed land-surface observations, CO₂ and water flux measurements, hydrograph data and remotely sensed CO₂ concentrations—to generate a time-varying field of gross primary production (GPP, the most fundamental function of all ecosystems) across the Australian continent.
<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Name</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macquarie University</td>
<td>Metadata Stores (MUMS)</td>
<td>The Macquarie University Metadata Stores initiative will aggregate the institutional metadata information for research collections. As an aggregator the system aims to interconnect data collection, which have native storage and provide the over-arching meta-information to facilitate discovery and usage of research information.</td>
</tr>
<tr>
<td>Monash University</td>
<td>Seeding the Commons Project</td>
<td>The Research Data Collections Project led by Monash University Library aimed to identify and describe research data collections arising from publicly funded research, and to showcase these collections by contributing information about them to Research Data Australia (RDA).</td>
</tr>
<tr>
<td>Research Data Management</td>
<td>of the Monash Weather &amp; Climate Program (Climate and Weather)</td>
<td>The Climate and Weather Data Capture Project addresses many issues facing Climate and Weather researchers in Monash University, supporting existing research and enabling Monash University Weather &amp; Climate (MW&amp;C) researchers to contribute to the work of evaluating the newly deployed Australian ACCESS climate model. Through ANDS funding and support the climate and weather research data has been transformed in the following ways: Overall, the Climate and Weather Data Capture Solution provides the MW&amp;C researchers with the ability to conduct more effective research based on exploiting existing data sets, creation and storage of new data sets and better collaborative research opportunities.</td>
</tr>
<tr>
<td>Biomedical Data Platform</td>
<td>(Molecular Biology)</td>
<td>The Biomedical Data Platform Data Capture solution, myTardis, is a multi-institutional collaborative venture that facilitates the archiving and sharing of data and metadata collected at major facilities such as the Australian Synchrotron and the Australian Nuclear Science and Technology Organisation (ANSTO) and within Monash University and other institutions.</td>
</tr>
</tbody>
</table>
| Tools for curating and    |                                                                             | The purpose of this project was to develop generic software tools to support the organisation of pre-existing digital data
<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Name</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>research data in the form of media collections (Multimedia Collections &amp; ARROW)</td>
<td>collections, particularly photographic data, into files and formats suitable for publication into institutional repositories and into other discovery services including Australian Research Data Commons (ARDC).</td>
</tr>
<tr>
<td></td>
<td>Capture and publication of Australian ecosystem data from a network of measurement sites (Ecosystem Measurements)</td>
<td>The solution provided by the Ecosystem project has enhanced the research process and provided new research opportunities. Overall, the solution cuts down the time significantly before the researcher can begin their research work and facilitates much simpler sharing of data. Due to these benefits the Ecosystem solution has been very well received by the OzFlux community of researchers. The next key challenge is to enhance the Ecosystem solution to facilitate Ecosystem research on a larger scale at the national level.</td>
</tr>
<tr>
<td></td>
<td>Capture and publication of data on the history of adoption (History of Adoption)</td>
<td>The History of Adoption Data Capture solution has provided automation of the capturing stories and associated metadata from the website submission form and storing them in a Data Management system; generating a story web page with attached story transcript, metadata files in MODS and DC formats and search tags; and &quot;publishing&quot; a story to ARROW, the Monash Library public repository and then into RDA.</td>
</tr>
<tr>
<td></td>
<td>Data Publication to Interferome (MIMR/Interferome)</td>
<td>The researchers at the Monash Institute of Medical Research (MIMR) use Agilent, a microarray platform, and large genomics data repositories like ArrayExpress and GEO. They require the ability to assimilate the research data from these complementary data sources into a single repository to facilitate comprehensive analysis. The system currently in place at MIMR does not capture research data directly from researchers and it also has limited capabilities to query the research data. The Interferome System addresses the issues of access, sharing, publishing and reuse, supporting existing research and enabling MIMR researchers to perform complete analysis and maximise the value of available research data.</td>
</tr>
<tr>
<td></td>
<td>Multimodal Kidney Imaging</td>
<td>This project will enable high-throughput, multi-modal, multi-resolution characterisation of ex-vivo animal model</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Murdoch University</td>
<td>Integrating precision agriculture</td>
<td>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.</td>
</tr>
<tr>
<td>Queensland University of Technology</td>
<td>Seeding the Commons Funding</td>
<td>Project was to identify and described QUT 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.</td>
</tr>
<tr>
<td>Greenhouse Gas Emissions from Australian Soils</td>
<td></td>
<td>The N2O Project aimed to re-engineer an existing client application called Morpho used in the ecological research domain. The Morpho software is used by researchers as part of work associated with the national N2O Network research program. Soil emissions data is automatically collected at various sites around Australia using automated gas sampling systems. A Java-based data management application called</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>MetaMaster</td>
<td>MetaMaster was developed to facilitate the management of Scientific Data Packages within the ecological research space and assists with the creation of metadata records that conform to the RDA metadata standard (RIF-CS). These are ingested into the QUT Metadata Hub published to Research Data Australia (RDA).</td>
</tr>
<tr>
<td>Biodiversity</td>
<td>Bio-diversity Data Capture Project</td>
<td>The Bio-diversity Data Capture Project aimed to study the management and publishing of audio data collected from acoustic sensors for environmental health. The researchers have developed a research application, the Environmental Health Sensor Monitoring (EHSM) system, which adopts acoustic sensing, database and web service technologies to capture audio data from different environments for monitoring species that have regular and predictable vocalisations. To help researchers to publish the information that is collected using the EHSM system, a Java-based desktop application called “RDB2RDF Mapper” was developed. The application assists a researcher and data manager to create a mapping that can then be used to transform the relational database data into a which can then be easily ingested into QUT’s Metadata Hub and published to Research Data Australia.</td>
</tr>
<tr>
<td>B150 Big Jam</td>
<td>Q150 Data Capture Project</td>
<td>The Q150 Data Capture Project aimed to study the management of metadata relating to multimedia data captured during Q150 Big Jam Live Music Festival. The multimedia data includes video, image and text that was recorded relating to the music, bands, artists and other background information during the festival. A Java-based desktop software application, “Media Crawler”, was developed to provide a view onto the file-system containing the multimedia assets, and allow the user to annotate, search and group the recorded assets. Metadata is automatically extracted from various file formats and other metadata can be imported. The software is meant to facilitate this process so that views, or “collections”, may be defined and exported with the associated metadata.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>A Data Transformation and Model Calibration System for Carbon and Nitrogen Dynamics in Australian Ecosystems</td>
<td>This project proposes to build a data transformation and model calibration system that will automatically or semi-automatically extract and transform historic data from multiple Australian sites and use them to calibrate and validate the predictive power of multiple carbon and nitrogen dynamics models. The software will capture the process in scientific workflow which will form the blueprint for future similar activities to calibrate and compare ecosystem models using Australian data.</td>
<td></td>
</tr>
<tr>
<td>QUT Research Data Finder</td>
<td>With the Metadata Stores Project funded by ANDS, QUT High Performance Computing and the Research and the Library will further implement and develop the VIVO system to become the QUT Metadata Registry system to be called QUT Research Data Finder. Full scale creation, management and workflow of metadata records which will be fully operationalised.</td>
<td></td>
</tr>
<tr>
<td>RMIT University</td>
<td>Screen Media Research Archive</td>
<td>The Media Virtual Research Collections (MaVReC) is a virtual collections service designed to integrate cognate but differently structured existing research collections into a responsive, sustainable, user-friendly application for the ingestion, search and retrieval of screen media research. Screen media research is uniquely characterized by a wide range of objects and outputs, Cataloguing and archiving each of these research outputs entails different data standards and management protocols and has resulted in disaggregation of research knowledge in the sector. By creating opportunities to interrogate assets and metadata from different collection frameworks, MaVReC aims to bridge the divide between the practice based screen media research (ERA Creative Works) and other forms of research.</td>
</tr>
<tr>
<td>RMIT University</td>
<td>Data Capture from High Performance Computing Multi-User Environments</td>
<td>RMIT is currently a very large user of state and national high-performance computing (HPC) facilities. This project will develop and deploy software tools and applications which will be deployed for the NCI Supercomputer National Facility, the VPAC Supercomputer Facility and the RMIT HPC Facility.</td>
</tr>
<tr>
<td>RMIT University</td>
<td>Online decision support toolkit for</td>
<td>This project will develop an online decision support toolkit which will enable Australian port authorities to make better decisions.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>climate resilient seaports</td>
<td>informed climate risk management decisions. This will involve three discrete components: the sourcing, refinement, and standardisation of multiple data sets needed for context specific climate change adaptation decision-making, the integration of both primary and transformed data (as produced by new extensible models being developed for the NCCARF project) within a suitable data management framework, and the development of innovative software that provides an interactive interface (developed in close consultation with port authorities) for considering future climate change impacts, the implications for organisational risk management strategies, and the identification of possible adaptation options.</td>
<td></td>
</tr>
<tr>
<td>Vincula</td>
<td>The focus of this ANDS-funded component will be to: create a central metadata store, provide for populating this with metadata from the implemented domain-specific metadata stores, provide a framework for the inclusion of new domain specific metadata stores, and to provide a mechanism to publish metadata to the ARDC.</td>
<td></td>
</tr>
<tr>
<td>Southern Cross University</td>
<td>Data Management Plan and Policy</td>
<td>Create an agreed Data Management Plan and Policy at SCU</td>
</tr>
<tr>
<td>Swinburne University of Technology</td>
<td>Swinburne’s “Watering the garden for the seeds to grow” project</td>
<td>This project has developed the framework for a data management service at Swinburne that is both customer-focused and suited to their local environment and resources. Discussions with researchers informed the development of both of our public-facing deliverables, the Research data collections survey and the Swinburne research data management checklist. Wherever possible work has been shared at <a href="http://www.swinburne.edu.au/lib/researchdata">http://www.swinburne.edu.au/lib/researchdata</a>.</td>
</tr>
<tr>
<td>Swinburne Metadata Store Project</td>
<td>The project will implement a metadata store at Swinburne. The metadata store will manage the description of research data held at Swinburne, with a number of benefits: Enabling improved compliance with funder requirements for research data management Increasing visibility around research data activity within Swinburne to enable strategic planning</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Enabling the discovery and reuse of quality Swinburne datasets, contributing to the research community as a whole Enabling high quality metadata to facilitate discovery and ongoing management.</td>
<td></td>
</tr>
<tr>
<td>University of Adelaide</td>
<td>Automated capture and publishing of data generated on high throughput plant phenomic platforms.</td>
<td>Extract and curate data generated by the Plant Accelerator's LemnaTec Smarthouse platforms. Publish the data to a navigable public repository, and metadata to Research Data Australia.</td>
</tr>
<tr>
<td></td>
<td>Genomics Data Capture</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td></td>
<td>Research Data Storage and Management</td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td></td>
<td>Soils to Satellites</td>
<td>Combine data from multiple study plots with a range of data “from soils to satellites” in an interface that allows for additional data exploration and later interpretation by researchers.</td>
</tr>
<tr>
<td></td>
<td>Research Metadata Store Project</td>
<td>This project aims to develop an institutional research metadata store to facilitate the collection, management, discovery and reuse of research datasets generated in the course of research activities at the University of Adelaide. The scope of the project includes definition of requirements, software development, installation, configuration and testing of software. The metadata store solution will utilise an existing open source solution. The project will leverage off work already being undertaken for the University of Adelaide’s Seeding the Commons project.</td>
</tr>
<tr>
<td>University of</td>
<td>Data Management</td>
<td>Create an agreed Data Management Plan and Policy at BU</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ballarat</td>
<td>Plan and Policy</td>
<td></td>
</tr>
<tr>
<td>University of Canberra</td>
<td>Cross-communication and enhanced accessibility for research data management systems</td>
<td>Create an agreed Data Management Plan and Policy at the University of Canberra</td>
</tr>
<tr>
<td>University of Melbourne</td>
<td>Seeding the Commons</td>
<td>This project was an eclectic exercise in spreading the word about the importance of research data, its management, its discovery, and its potential re-use. Through the exercises undertaken they have been able to communicate to a broad range of individuals and groups across the university. They have partnered with key stakeholders and developed strong relationships that will hold them in good stead for continuing the work that has been started and built upon.</td>
</tr>
<tr>
<td>Enhanced Metadata Capture for Sustainable Management, Sharing and Re-use of APN Histopathology Research Data</td>
<td>This project helped establish a national database of mouse pathology to enhance the utilisation of mouse models of disease by Australian researchers. It enhanced metadata capture facilities for the Histopathology and Organ Pathology Service based at the Department of Anatomy and Cell Biology, The University of Melbourne as part of facilitating the sharing and re-use of mouse pathology research data both now and into the future. The project addressed current metadata scalability and sustainability issues associated with the service in order for the Melbourne Histopathology Service to participate in and contribute to emerging research data networks like PODD and ANDS.</td>
<td></td>
</tr>
<tr>
<td>Melbourne Neuropsychiatry Centre (MNC) Bioinformatics Development Project</td>
<td>The MNC has one of the largest databases of brain scans and associated neuropsychiatric research data in the world. It has National and International collaborators using and contributing to the database. Broadly ANDS-related activities will involve:&lt;br&gt;- Building a workflow for automatic documentation of dataset segments used in individual studies and publications; and for automating creation of citable persistent identifiers for unique studies and linking with</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Building software to automate capture of public facing metadata to University of Melbourne Registry which will deliver collections metadata to the ARDC.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- MNC has 270+ publications resulting from datasets stored in the MNC database. Completing the work above will result in ~ 125 dataset descriptions described in the ARDC.</td>
</tr>
<tr>
<td>Youth Research Centre’s Life Patterns Project:</td>
<td></td>
<td>The Youth Research Centre’s Life Patterns Research Program maintains an extensive qualitative and quantitative data base on a cohort of 2000 young Australians who left secondary school in 1991 and of a second cohort of 3000 who left school in 2005. The project resulted in the implementation of a local system and related workflows and practices to describe the collection. Metadata about the Life Patterns Research Program is now exported to the University Research Data Registry and harvested to Research Data Australia.</td>
</tr>
<tr>
<td>Longitudinal qualitative and quantitative survey data capture and reuse</td>
<td></td>
<td>Video data in the Social Sciences. Optimising Metadata Capture, Data Sharing Procedures and Long-term Reuse</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The University of Melbourne has an especially rich humanities and social science research community that utilises video as its primary form of data capture. The increasing use of video as a research tool poses particular challenges for aggregated data storage initiatives. This project will integrate metadata capture facilities at selected sites within the University of Melbourne as part of facilitating sharing and re-use. The project will address current metadata issues associated with large-scale audio-visual repositories and workflows to enable efficient generation of metadata, ensuring that stored video data is accessible and searchable through the ARDC.</td>
</tr>
<tr>
<td>Federated Neuroimaging Collections in the National Data Commons</td>
<td></td>
<td>DaRIS is a raw data management system based on the Mediaflux digital asset management platform and has been in operation for the last 3 years at the Neuroimaging Computational and Data Management Facility (CDMF). There it has been used to routinely receive MR images from researchers and organise them into a subject-centric data model, ready for access by project members. It hosts over 70 mouse and human projects, each with many tens of subjects and some with time-dependent data.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Humanities and Social Science Research Data at the University of Melbourne</td>
<td>HASS researchers at Melbourne generate and hold valuable data sets and associated materials that are currently not easily discoverable, accessible or configured for further research purposes. This project will build infrastructure (tools and services) to connect this diverse community with the UoM Registry which will in turn communicate the relevant metadata to RDA.</td>
<td></td>
</tr>
<tr>
<td>Capture of Complex Data to Support Clinical Research in Cardiovascular and Neurological Medicine</td>
<td>Complex physiological data is routinely collected on patients as part of clinical. However, this rich multi-model data is not usually subjected to subsequent analysis nor is it made available to researchers from other disciplines for novel analysis. Making this multi-model data available along with patient outcomes such as morbidities will provide the opportunity for collaborative groups to employ novel strategies to developed assessments and models based on this data. This project will form necessary base of making multi-model data collections available, enabling the establishment of new links between biomedical research groups in engineering, physics and bioinformatics. This project will occur in collaboration with BioGrid Australia where it will use the access, de-identification and privacy protection protocols already established there.</td>
<td></td>
</tr>
</tbody>
</table>
| Founders and Survivors Project | The Founders and Survivors Project (http://www.foundersandsurvivors.org/) has brought together a number of research data sets created from records relating to the 73,000 convicts transported to Tasmania in the 19th century and their descendants to create a population database of significance for historical, demographic and population health researchers. This project has:  
- Developed a toolkit based around the projects XML/TEI workflow for further relevant records sets to be systematically ingested into the population database,  
- Built the infrastructure to enable persistent identification and descriptions of derived data sets produced on request from the population database to be made available to the ARDC |
<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Name</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Founders and Survivors:</td>
<td>The Founders and Survivors Project</td>
<td>The Founders and Survivors Project (<a href="http://www.foundersandsurvivors.org">www.foundersandsurvivors.org</a>) has brought together a number of research data sets created from records relating to the 73,000 convicts transported to Tasmania in the nineteenth century and their descendants to create a population database of national and international significance for historical, demographic and population health researchers. This project will develop open-source software infrastructure based on Yggdrasil (solumslekt.org/forays/yggdrasil.php).</td>
</tr>
<tr>
<td>Genealogical Connections:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AURIN &amp; ANDS – North West Metropolitan Region of Melbourne Data Access, Integration and Interrogation and Demonstrator Projects</td>
<td>The purpose of this project is to facilitate access to a myriad of data sets for the Melbourne North West corridor. The aim of the project is to demonstrate the benefit of providing open access of government datasets to researchers, planners and policy makers in dealing with problems of space, place, and liveability. The value of the project will be demonstrated through four demonstrator projects which cover four of the most pressing issues facing the North West Melbourne Region: built environment and health, housing affordability, economic productivity, and transport and sustainability.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Data Registry</td>
<td>This program will develop metadata store infrastructure to manage metadata about data collections, activities and parties, and to provide feeds of this metadata in the form of RIF-CS XML</td>
<td></td>
</tr>
<tr>
<td>University of New England</td>
<td>UNE’s N.C.W. Beadle Herbarium Database</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of New South Wales</td>
<td>Research Data management Services</td>
<td>The Seeding the Commons project at UNSW has developed a number of services, applications and resources to support research data management: more than 80 research collections were documented and contributed to Research Data Australia. Records of associated projects, services, people and organisations have also been made available; resources about research data management have been developed and disseminated; the ResData Deposit Tool was developed to enable UNSW researchers to contribute.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>An international antibiotic-resistance gene cassette database</td>
<td>The project made technologies that allow us to archive relevant elements and to identify them in bacterial DNA sequences. It has also built a knowledge repository of antibiotic resistance gene cassettes available to the wider research community and to allow the community to contribute new entries to the repository as these are found. Providing this application should enhance the team’s global reputation as leading researchers of antibiotics resistance and pioneers in the analysis of larger-than-gene structures.</td>
</tr>
<tr>
<td></td>
<td>Managing and Sharing Genomic Data</td>
<td>A new generation of DNA sequencers has recently been installed at UNSW, Southern Cross University and the Australian National University. These instruments can generate DNA sequence data 1000x faster than old technology, and can sequence the genomes of small organisms in a week. This project established databases for this DNA sequence information, so that users of the DNA sequencers can access their information in an efficient way.</td>
</tr>
<tr>
<td></td>
<td>Australian and New Zealand Neonatal Network (ANZNN) Neontal Data Capture Portal</td>
<td>Building software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td></td>
<td>Data capture and integration across multiple platforms</td>
<td>Building software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td></td>
<td>Validation of genomes and</td>
<td>This project will integrate genomic, transcriptomic and proteomic data to validate open reading frames in genomes.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>transcriptomes with proteomic data</td>
<td>It will also use proteomic data to validate isotigs (mRNA splice variants). Results will be tabulated and available for browsing in a genome viewer.</td>
</tr>
<tr>
<td></td>
<td>Climate Model Downscaling Data For Impacts And Adaptation Research</td>
<td>Climate impacts research depends on the timely provision of data sets that describe how climate will change across a region. These data sets are created by climate scientists in a discipline specific format via a downscaling procedure designed by climate scientists for climate scientists. This ANDS project will build tools that identify suitable downscaled data sets and read the contents of those data sets. A dialog with four key impacts research groups will be established to define their data needs for their research. The tools will be developed to generate these data, as required by the impacts researchers. The data will be exported in formats defined by the impacts researchers to enable their research. All tools, and all generated data sets will be made open source, hosted on an RDSI node and be citeable.</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>UNSW Metadata Stores Project</td>
<td>The UNSW Metadata Store project will deliver an enterprise metadata storage system.</td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>Newcastle Research Data Online (Seeding the Commons)</td>
<td>The University of Newcastle collaborated with a team from the University of Southern Queensland on the development and deployment of ReDBox 1.0 at the University of Newcastle. Web based ingest forms were developed to capture rich metadata around research data collections. The project implemented strategic systems, interfaces and processes to enable the capture of metadata. It also identified and developed interfaces or linkages to predetermined strategic triggers within the organisation to assist with the identification of potential relevant metadata capture. The project also focused on populating NOVA, the institutional research repository, with Research Data Collection records including links to research data wherever possible.</td>
</tr>
</tbody>
</table>
|                      | Data Capture for the Data Commons                                             | Building software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be...
<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Name</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newcastle Research Data Online: Stage 2</td>
<td></td>
<td>connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td>Newcastle Research Data Online: Stage 2</td>
<td>The Newcastle Research Data Online Project: Stage2 adheres to the ANDS funding requirements by building on the foundation of our existing ReDBox/Mint metadata store solution. The project will enhance functionality to meet mandatory deliverables including RIF-CS 1.3 compliance, and additional institutional triggers via development and system upgrade. The project will investigate internal data flow and data sources in current corporate, research and IT systems to inform efficient and effective systems integration for ReDBox/Mint at Newcastle. The project will identify and potentially develop software requirements at an institutional level to build connections between ReDBox/Mint and internal sources of truth via systems integration. This will create the linkages required for automatic system feeds to provide and maintain institutional records for Parties, Collections and Activities from ReDBox/Mint.</td>
<td></td>
</tr>
<tr>
<td>University of Queensland</td>
<td>Seeding the Commons Funding</td>
<td>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 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</td>
</tr>
<tr>
<td>Spatially Integrated Social Science</td>
<td>The Urban and Regional Analysis Research Program (URARP) in the Institute for Social Science Research (ISSR) is developing a repository for capture, integration and sharing of diverse socio-spatial data sets.</td>
<td></td>
</tr>
<tr>
<td>Aquatic Species Tracking Repository</td>
<td>This project is collecting data captured by the ECO-Lab at the University of Queensland (Prof Craig Franklin and Dr Hamish Campbell) - who are using arrays of underwater,</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>acoustic receivers to track aquatic animals (crocodiles, turtles, rays and sharks) within river systems over large temporal (3-10 years) and spatial scales (100's km). The data is being collected to improve our understanding of the ecology and habitats of these species and to provide information to aid in their conservation and management.</td>
<td>The Health-e-Reef Project This project is developing the data capture and sharing services for coral-reef related data being generated by researchers at the University of Qld Centre for Marine Studies, together with their collaborators in community/volunteer groups (CoralWatch, ReefCheck) and government organizations (EPA, DERM). Together these researchers are monitoring and studying the impact of climate change and human activities on coral reef ecosystems. There will be particular focus on automatically capturing the metadata necessary to support discovery, decision, and reuse. The discovery metadata will be made available through RDA.</td>
</tr>
<tr>
<td>Microscopy/Microanalysis Image and Data Repository</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
<td>The Diffraction Image Experiment Repository (DIMER) 3D Anthropological and Archaeological Collection Repository Linking the EMBL Australia EBI Mirror with the Australian Research Data Commons Building software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>compare and quantify differences between multiple biological specimens (e.g. wild-type vs. disease specimens). During the course of the project the system will cater for adult mouse brain imaging MRI and histology data, but will be built in a species independent manner.</td>
</tr>
<tr>
<td>Cancer Genomics Linkage</td>
<td>The Cancer Genomics Linkage Application project will develop software that will enable researchers to access, integrate and transform unique and valuable cancer genomics datasets from the International Cancer Genome Consortium and other sources and to analyse them using the computational capability of the Federally supported Genomics Virtual Laboratory. The integrated data will afford the cancer genomics researcher the opportunity to more rapidly investigate their own data in light of the public data in a comprehensive and re-useable manner.</td>
<td></td>
</tr>
<tr>
<td>UQ Data Collections Registry</td>
<td>This project will develop, test, configure and deploy a UQ metadata store for research data collections that aggregates metadata from existing sources of truth and feeds metadata to ANDS Research Data Australia, the ARDC Party Infrastructure, and the UQ Operational Data Store. The project will also develop a specification of portal requirements for a UQ Academic Portal, and analyse potential software solutions meeting those requirements.</td>
<td></td>
</tr>
<tr>
<td>University of South Australia</td>
<td>Taking Australian Architectural and Built Environment Records into the Commons</td>
<td>Metadata on collections held at the Architecture Museum was originally available in PDF Finding aids on the museum’s website. All collection and party metadata has been entered into Metatecture and can be searched by UniSA staff and the public using the Public Interface to Metatecture. Metadata is now stored in a consistent manner in a centralised location. Metadata is also available to a wider audience via Research Data Australia (RDA), Trove etc.</td>
</tr>
<tr>
<td>Development And Testing Of A Data Capture Tool For Social Datasets Being Used For Record Linkage</td>
<td>The Ian Wark Research Institute (IWRI) is a major node for both the Characterisation and Fabrication initiatives, funded by NCRIS with supplementary EIF support. Both initiatives feature unique (in Australia) flagship instrumentation together with IWRI’s existing equipment infrastructure, which generate data of a variety of types and sizes. The project is developing a MetaData Capture Tool, a semi-</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>UniSA – Research Metadata Store</td>
<td>automatic tool to facilitate capture of both data and metadata from the IWRI TRIFT V and Mastersizer 2000 instruments.</td>
<td>UniSA’s project will implement an enterprise metadata store solution. The system will be developed as an enterprise class metadata store that will be an enabler of complete, high quality, compliant and accessible metadata for UniSA’s research activities.</td>
</tr>
<tr>
<td>University of Southern Queensland</td>
<td>Sustainable policy and procedure for capturing research data</td>
<td>Create an agreed Data Management Plan and Policy at USC</td>
</tr>
<tr>
<td>University of Sydney</td>
<td>Clarke eHealth (Early Activity): Capture, management, re-use and discovery of breast cancer microscopy virtual images</td>
<td>This project constructed software to allow wide use and re-use of microscopy images in breast cancer research, generated at the Westmead Institute for Cancer Research (WICR), in analysis and study by approved researchers across Australia, with collection descriptions available through RDA.</td>
</tr>
<tr>
<td></td>
<td>AgDataCapt: Capturing Agricultural Data</td>
<td>Standards-based body of data and metadata will be coordinated and include data from sensors on soil, water and rain, greenhouse gas and carbon data, and weather. Generic and extensible tools will be developed to integrate appropriate standards across the areas of data capture. The areas of data capture are: Soil moisture, soil and air temperature, radiation, 3D wind speed and directions, CO2, water vapour, weather and tree water use data, greenhouse gas emissions from soils; GPS-referenced data on wheat and barley crop inputs and performance; Soil data from sensing system for monitoring of agro-ecosystems for sustainable landscape; Soil data and spatial prediction functions for soil variables; and Data from farm-based private rain gauges.</td>
</tr>
<tr>
<td></td>
<td>AMMRF Live Cell Microscope Data Capture</td>
<td>Experiments conducted using high-end microscopy instruments currently record information in multiple dimensions, such as 3D and 4D, often with multiple detectors. Research data collections are therefore increasing in size as microscopy techniques evolve, often</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>resulting in extremely large images that are difficult to manage using current infrastructure and systems. Images also often require post-acquisition processing steps where the research data moves between analytical platforms, which can be situated in different sites. The University of Sydney DC2E Microscopy data capture project will complement the NeAT Australian Microscopy and Microanalysis Research Facility (AMMRF) PfC project by catering for additional light microscopy instruments, extending the scope for the repository that the current project is developing.</td>
</tr>
<tr>
<td>Marine video storage and analysis</td>
<td></td>
<td>This project will provide funding to support the development of a database framework applicable to marine imagery that will facilitate the timely delivery of standardized, quantitative estimates of ecologically relevant indicators (such as absence/presence, percent cover, abundance and distribution of benthic organisms and associated substrates) through the analysis of visual data being produced by AUV, BRUVS, ROV and UTV systems.</td>
</tr>
<tr>
<td>NSW TARDIS Node; FieldHelper: a workflow and tools for improving fieldwork data collection and submission to institutional repositories</td>
<td></td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td>Seeding the Commons at the University of Sydney</td>
<td></td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td>Brain and Mind Research Institute</td>
<td></td>
<td>This project will support new and existing multidisciplinary research through the integration of research image scans and clinical data overlaid with analytical tools. The project will also directly transfer benefits of research to the clinical domain. Findings from research scans will be returned to referring clinicians, providing additional diagnostic support.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>University of Sydney</td>
<td>University of Sydney Metadata Store</td>
<td>This project will provide an enterprise-wide research data management solution by bringing together existing components, filling gaps, providing a metadata store for research data collections and a system for managing metadata created through StC. In support of strategic planning, the system will enable reporting on published collections. Metadata management for unpublished collections will not be within scope, but will be recommended for attention post-project.</td>
</tr>
<tr>
<td>University of the Sunshine Coast</td>
<td>Data Management Plan and Policy</td>
<td>Create an agreed Data Management Plan and Policy at USC</td>
</tr>
<tr>
<td>University of Tasmania (UTAS)</td>
<td>Publication of collections into the ARDC by UTAS</td>
<td>The Tasmanian Partnership for Advanced Computing (TPAC) hosts and publishes an extensive range of data sets from around 150 separate collections. The collections hosted and/or published by TPAC comprise data from approximately one million NetCDF data sets, and total approximately 50 terabytes (at TPAC) and a further 30 Terrabytes as part of this federation of Open-source Project for a Network Data Access Protocol (OpenNDAP) servers at CSIRO and ANU <a href="http://dl.tpac.org.au/">http://dl.tpac.org.au/</a>. The collections are produced by research organisations both internal and external to UTAS.</td>
</tr>
<tr>
<td></td>
<td>Data Capture of state-wide hydrological datasets</td>
<td>The aim of the project is to capture and publish data from the CSIRO and Forestry Tasmania sensor webs. The data will be exposed by two sensor observation services (SOS) serving twenty two sensors, and a THREDDS server providing forecast data.</td>
</tr>
<tr>
<td></td>
<td>Redmap Australia</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td></td>
<td>Tasmanian</td>
<td>The project proposes to create a Tasmanian Research eData</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>University of Technology Sydney (UTS)</td>
<td>Research eData Directory Service (TReDDS)</td>
<td>Directory Service which describes key elements of a diverse range of data stores held by Tasmanian research entities. This directory service will interface to national metadata stores and hence provide a more extensive and coherent discovery service to data assets held within the state. Management of the service will be coordinated via the UTAS library, through curation, archival, harvesting, and publishing frameworks.</td>
</tr>
<tr>
<td>University of Technology Sydney (UTS)</td>
<td>Community Tools and Processes for Effective Data Management Planning</td>
<td>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.</td>
</tr>
<tr>
<td>University of Technology Sydney (UTS)</td>
<td>Maximising the Benefit from Data-Intensive Processes at UTS</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td>University of Technology Sydney (UTS)</td>
<td>UTS Metadata Stores Project</td>
<td>Selection, testing and installation of an enterprise metadata store. Integration of this metadata store with the University's existing research data systems and with data capture activities funded by ANDS Data Capture project DC18.</td>
</tr>
<tr>
<td>University of Western Australia</td>
<td>Deployment and configuration of Institutional Metadata Repository</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td>University of Western Australia</td>
<td>Integrated Data Capture for Characterization and Analysis</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Archaeological Rock Art Data Capture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marine Ecology Video Capture and Storage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seeding the Commons through research data management at the University of Western Australia</td>
<td>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</td>
</tr>
<tr>
<td></td>
<td>POSITIVE PLACES: spatial analysis of public open space</td>
<td>This project aims to integrate heterogeneous data sets from POS data layers with other geo-spatial build environment information and demographic and socio-economic status data, and to develop a web based geospatial data analytic and visualisation tool that would support two new opportunities in research and for urban planning practice.</td>
</tr>
<tr>
<td></td>
<td>UWA Institutional Metadata Repository</td>
<td>This project will build on the metadata store (VIVO) implemented by UWA under the ANDS DC9 project (see Deployment and configuration of Institutional Metadata Repository above). This project (“Stage 2”) will expand the scope of the UWA metadata store to include Higher Degree by Research students and add information about research publications derived from data collections. It will also ensure that the UWA metadata store is more closely aligned with external sources of truth (e.g. ARC and NHRMC grant registries, and NLA Party identifiers) and will investigate ways of increasing the alignment between the metadata store and data management workflows.</td>
</tr>
<tr>
<td>University of Western Sydney</td>
<td>Climate Change and Energy Research Facilities (CCERF)</td>
<td>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</td>
</tr>
<tr>
<td>Institution</td>
<td>Project Name</td>
<td>Project Description</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>UWS Seeding the Commons</td>
<td>Working with the Data Capture projects funded by ANDS, to apply the lessons learnt and help create a broader data management policy framework.</td>
<td></td>
</tr>
<tr>
<td>Enterprise Research Data Catalogue</td>
<td>Development of a research data catalogue containing metadata about data at a collection level for code-compliance, strategic research management and discovery purposes.</td>
<td></td>
</tr>
<tr>
<td>University of Wollongong (UOW)</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 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. Develop university wide data management procedures relevant to UOW.</td>
<td></td>
</tr>
<tr>
<td>Biomechanics Data Capture Project System</td>
<td>This project aims to: create metadata entries associated with datasets. The tagging of metadata will be a combination of automatically from the devices and manually entered by the user; provide a central storage repository where metadata coupled with raw datasets can be stored; provide a mechanism whereby collections of data can be defined; provide a mechanism whereby metadata can be harvested and automatically transported to RDA; and provide researchers with an interface to search for and download datasets associated with their research.</td>
<td></td>
</tr>
<tr>
<td>SMART’s Multi-Utility Dashboard - Infrastructure Analytics for Integrative Research</td>
<td>The SMART Infrastructure Facility will develop a ‘multi-utility dashboard’ that will offer infrastructure analytics based on data provided by public agencies and private operators. The online dashboard will allow analysts to develop new insights into spatial, technical, social and economic issues associated with regional and urban infrastructure development. The multi-utility dashboard will act as ‘one-stop-shop’ portal accessing, formatting, analysing and making publicly available information on water, energy, waste, communication and transport distribution or management</td>
<td></td>
</tr>
</tbody>
</table>
networks in a given area. This information, crossed with relevant figures from demographics and economics will constitute a robust foundation for powerful infrastructure analytics. The multi-utility dashboard will offer a flexible collaborative platform to researchers, business analysts and local planners.

Satellite data capture

Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Name</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>networks in a given area. This information, crossed with relevant figures from demographics and economics will constitute a robust foundation for powerful infrastructure analytics. The multi-utility dashboard will offer a flexible collaborative platform to researchers, business analysts and local planners.</strong></td>
</tr>
<tr>
<td></td>
<td>Satellite data capture</td>
<td><strong>Building infrastructure between instruments and well supported data and metadata storage facilities; software to enable better management and descriptions of research data and associated metadata; and use of this infrastructure to better manage and describe research data and associated metadata to enable it to be connected, and feeding the relevant records into the ARDC to facilitate discovery of the data and then re-used.</strong></td>
</tr>
<tr>
<td>Victoria University</td>
<td>Research data framework</td>
<td><strong>A wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy.</strong></td>
</tr>
</tbody>
</table>

### 4.3.4 2013-14 Activities

The scope of this program is data-focused engagement with all of Australia’s major research organisations. This will carry the foundation projects forward into assisting the institutions moving into a business as usual environment for the delivery and support for data management and publication.

Specific activities under this program include:

**Provide targeted assistance to selected institutions to enable routine publication of research data outputs conformant to national guidelines**

The basis for this targeted assistance will be the development of a framework for data publication which will incorporate the following elements: Discoverability, Accessibility, Usability, Citability. A defined ANDS service offering will include advice, tools and training that can be offered within a set of categories such as management, publication citation, re-use.

In order to deliver the service appropriately the program will undertake a skills inventory in ANDS to determine what skills ANDS has to support the service offering. It will also compile an inventory of business requirements from research institutions and data providers and funders to identify the projects they are funding and their requirements for data publication. This latter has the aim of identifying some of the major funders with whom ANDS can work to publish data that is an outcome of the research they fund.
Support institutional engagement approaches that enable better sharing, improved efficiency and greater coherence of research institutional approaches to all institutional research data infrastructure

This activity will continue and improve current practice through the development of a community of practice in data management, publication and re-use and of content and resources to support communities and training needs. It will involve a review of current relationships and communities for ongoing maintenance and management and the enhancement of facilities to act as a dissemination point for useful resources. This will involve an audit of institutional business requirements to identify training and community action needs and the development of an inventory of types of approaches we can employ (both ANDS instigated and externally instigated that we can join) and when they would be applicable. These could include community days, webinars, surgeries, blogs and more. This outreach will be extended to include engagement with research forums that are receptive to ANDS participation.

Facilitate the re-deployment of data tools in new environments

A review of the current product catalogue will be undertaken to align the presentation of the tools to ANDS service offerings. This will involve the identification of any relevant tools for service categories and the source of support for those tools. Deployment facilitation will also involve the development of a protocol for assistance offered – referral, technical support etc. This will vary according to tool; an initial heuristic could be that technical assistance is sought in the first instance from the source of the code, for example: Orca = ANDS; embargo tool = CSIRO, ReDbox = QCIF.

This activity will also include the facilitation of a community of practice in tools reuse to share business requirements and experience of tools uptake among the research data community. It will also involve work with NeCTAR and RDSI to facilitate and demonstrate tools redeployment involving their environments and the promotion of cloud deployments of tools through international forums.

Demonstrate the value of better use of research data through the Application program and associated activities

Over the life of the Applications program examples of good data re-use practice (as well as associated data combination and analytic tools) will have been developed to demonstrate the value of bringing data together to answer new questions. The software will have been developed to meet the needs of particular research groups, but may have potential for adaptation and redeployment in other institutional settings. The instances of data re-use good practice will be repurposed by the Institutional Engagement program as exemplars for continued data management services. These exemplars can be promulgated through a variety of mechanisms within the community of practices.

4.3.5 Highlights

A highlight as evidenced by Table 6 outlining the infrastructure developed to date has been the extent of the reach of the projects that ANDS has funded. Data management has been placed on the
agenda on the vast majority of Australia’s research institutions. The extent of this reach has also ranged across discipline boundaries and in a number of cases, solutions for data management has been cross disciplinary which has set a foundation for further integration into the future.

The development and uptake of tools has also been a highlight especially in the area of metadata stores solutions. There has been a willingness of a number of institutions to collaborate to reduce duplication of effort and implement shared and innovative solutions across organisational boundaries. This has been a characteristic of the growing community of practice which is developing a body of expertise supporting Australian research data.

Within institutions there has been the development of effective partnerships to support data management, publication and re-use. These partnerships involve eResearch, IT, Libraries, Research offices and research groups. An outcome of dialogue between these groups is an increasing awareness around the broader context for data and the need for interoperability between the organisation’s systems to effectively support this.

### 4.3.6 Challenges

One of the challenges that has emerged in the last period has been one of scheduling. Meeting contract milestones and completing projects within the ANDS timeframe has proved a challenge for a number of projects. A small number of projects will thus continue for a short period into the 2013-14 planning timeframe.

A challenge for future sustainability will be resourcing the services within organisations without the additional injection of funds that was provided through the ANDS contracts. This will be especially critical in a new funding environment that is promising to be very tight.

ANDS service development will require a cultural change within ANDS as staff move from overseeing the completion of contracted projects to engaging more closely with colleagues in institutions. This will also precipitate a change in the relationship between the ANDS staff and their contacts in institutions who will have developed a body of on-the-ground practical experience; their requirement for assistance will therefore necessarily be more sophisticated. The environment is now a complex network and partnerships with research institutions will need to be maintained in conjunction with relationships with other eResearch infrastructure partners such as RDSI and the nodes and NeCTAR. Cooperative engagement will be vital for the success of ANDS’ institutional engagement.

The projects have delivered a number of software solutions, some of which have had uptake from other institutions. An ongoing challenge is how best to support ongoing development and deployments of the software solutions that have been created now that this program has committed all its funds.

### 4.3.7 End of 2013-14 Outcomes

By the end of 2013-14 ANDS will have:
Placed Data management on the agenda on the vast majority of Australia’s research institutions

Assisted institutions to set up the infrastructure to support their research data ambitions and the notion of data as an institutional asset

Underpinned the ability of institutions to contribute to distributed national collections on topics of national significance

Developed an awareness of the importance of data management in a number of disciplines which is leading in turn to increased data exchange across discipline boundaries

Facilitated effective partnerships between eResearch, IT, Libraries, Research offices and research groups within institutions

Strengthened communities of practice for developers and research data managers

4.3.8 End of 2013-14 Enduring Outcomes

By the end of the funding period ANDS will have produced the following enduring changes through this program:

- A national resource of managed, connected, findable and re-usable research data now exists that previously did not.
- More data are being automatically captured from a wide range of instruments, ranging from small sensors to major national facilities.
- Data providers are increasingly seeing the value in providing their data through richer shared environments.
- Data are now visible to researchers that were never visible before.
- A cohort of leading Australian researchers has demonstrated the value of a richer data environment.
- Australian researchers can publish and cite their research data outputs.
- Larger questions and grander challenges can be addressed by a data commons that spans facilities, institutions, disciplines and sectors.
- New types of research are now possible that previously were not.
- Data can be captured, managed, and shared more easily at most research organisations in Australia.
- There is more cost-effective reuse of research data software in Australia.
- Research organisations and public sector agencies now provide policy support and invest more in data management.
- Australian research institutions can more easily comply with the requirements of the Code for the Responsible Conduct of Research.
- There is an established community of data management professionals.
- Australian researchers and research organisations have better capability to operate and exploit the new research data infrastructure.
- Common approaches, standards, and services have increased efficiency and coherence across the national data infrastructure investment.
- Established data management infrastructure now allows for stronger national policy and easier institutional compliance.
- Australia is seen to be a leader in research data infrastructure internationally.

### 4.4 International Collaboration

#### 4.4.1 Program Aims

This program aims to improve collaboration with data infrastructure providers inside and outside Australia to ensure that Australian research data infrastructure is compatible with international approaches.

#### 4.4.2 Program Overview

Until 2012, international collaboration activities within ANDS occurred in a number of different programs and arising from a combination of strategic decisions and consequences of preceding projects. The strategic decisions included the selection of DataCite as our preferred Digital Object Identifier solution, and engaging with international partners to adopt our technology as a way of future proofing the investment and expanding the pool of potential developers. An example of the latter is the Digital Curation Centre being asked by JISC to evaluate the Research Data Australia software as a collections registry solution for the UK. A number of ANDS staff had worked on preceding projects in the institutional repository, virtual research environment and e-research space and brought this experience and the associated personal and professional relationships built up over a number of years to discussions about international collaboration.

The First EU-AU Workshop on Research Infrastructure, held in Brussels in April 2011 brought about the start of co-ordinated activity that would eventually lead to the establishment of the Research Data Alliance. This was accelerated at the Second EU-AU Workshop on Research Infrastructure, held in Brussels in June 2012. The Research Data Alliance has now become the main vehicle for ANDS’ involvement in international activity, although a number of existing activities will continue in parallel. This means that from now on, potential international collaborations will first be evaluated in terms of how they might contribute to the Research Data Alliance agenda as part of our commitment to this activity.
The purpose of the Research Data Alliance is to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability. This will be achieved through the development and adoption of infrastructure, policy, practice, standards, and other deliverables. The Research Data Alliance has been funded by an initial three research funding organisations:

- The Australian Commonwealth Government through the Australian National Data Service, supported by the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative, and through a special contract, entitled the Data Web Forum, specifically targeted at ANDS’ involvement in the Research Data Alliance.
- The European Commission through the iCordi project funded under the 7th Framework Program
- The United States of America through the RDA/US activity funded by the National Science Foundation

The work of the Research Data Alliance will primarily be undertaken through its working groups and interest groups. The development of solutions will occur through working groups. These are intended to come into existence, work on a problem and deliver a solution that is both adopted and that improves data exchange at the end of a 12-18 month period. Interest groups are for people with an interest in a particular data technology or research discipline. Interest groups may identify data interchange problems that need to be solved. Participation in working groups and interest groups, starting new working groups, and attendance at the twice-yearly plenary meetings is open to all.

4.4.3 Infrastructure Created to Date

Not relevant to this program as yet.

4.4.4 2013-14 Activities

Work to establish and develop the Research Data Alliance

- Contribute to RD-A Organising Group meetings and associated activities in lead up to Launch
- Contribute to the RD-A Secretariat
- Take part in Council Meetings
- Develop policy and operations documents
- Liaise with Working Groups to assist them to engage with RD-A processes in lead up to Launch
Contribute to ongoing operations of Research Data Alliance

- Contribute to RD-A Organising Group meetings and associated activities
- Contribute to the RD-A Secretariat
- Take part in Council Meetings
- Contribute to other supporting bodies of the RD-A, as required
- Join and contribute to Interest and Working Groups, as required
- Develop and refine policy and operations documents
- Liaise with Working Groups and Interest Groups to assist them to engage with RD-A processes
- Take responsibility for the program of the Third Plenary to be held in March 2014
- Contribute to the process of setting up a legal entity for RD-A
- Facilitate involvement of other countries and institutions in RD-A

Facilitate Australian involvement in Research Data Alliance working groups

- Work with Linguistics community to develop proposal for ISOCat language codes as RD-A working group
- Work with Marine community through IMOS as they develop a proposal for a Marine RD-A working group
- Actively seek and work with other Australian representatives of other communities for whom an RD-A working group is an appropriate vehicle to develop and implement international data exchange technologies

Contribute to success and expansion of persistent identification of data through DataCite

- Attend DataCite Board Meetings and Annual General Meetings
- Promote DataCite as a technology for data citation

Contribute to success of data and research identification integration through ODIN

- Deliver against work packages assigned to ANDS in ODIN project plan

Align work plans between British and Dutch organisations, including the JISC-MRD program, DCC, and DANS

- Share ANDS work plans (including this document) with JISC Managing Research Data program (UK), Digital Curation Centre (UK), Data Archiving and Network Services (NL), SURF 9NL)
• Identify areas of common interest and plan how to co-ordinate activity through to end 2014

4.4.5 Highlights
The major highlight to date for activity under this program has been the involvement of Australia through ANDS in the formation of the Research Data Alliance. The invitation for Australia to join the research powerhouses of the USA and Europe in bringing about this new organisation is a very real validation of the investment that Australia has been making in research data infrastructure.

A related highlight was the launch of the Alliance in Gothenburg, Sweden in March 2013. This event, opened by Vice Commissioner Kroes from the EC, with Ambassador Lewis from Australia and Dr. Jahanian of the NSF, was viewed by all those who attended as very successful and contributed markedly to the momentum behind the Alliance.

Finally, the Executive Director of ANDS, Dr Ross Wilkinson, was chosen as one of the three inaugural members of the Research Data Alliance Council, the RD-A’s ultimate decision-making body.

4.4.6 Challenges
One of the difficulties in this program relates challenges of maintaining effective collaboration with overseas colleagues across time and cultural differences. Another difficulty is working out how best to maintain a balance between (currently) the US, EU, and AUS involvements in and contributions to RD-A, including managing expectations, particularly given very different funding levels.

A third challenge relates to engaging with disciplines in Australia that are already collaborating effectively international within their discipline, in a way that adds value and coherence to the sector as a whole.

4.4.7 End of 2013-14 Outcomes
By the end of the planning period, the following outcomes will have been delivered:

• The Research Data Alliance will have all of its governance components defined and in operation
• A number of Research Data Alliance Working Groups will have significant Australian involvement
• A second and third RD-A plenary will have been held.
• DataCite will be the default identifier for datasets
• Effective linking of research data and researchers will be demonstrated through ODIN
• Workplans will have been shared with JISC, the DCC and DANS

4.4.8 End of 2013-14 Enduring Changes
By the end of the planning period, the following enduring changes will be in place:
The Research Data Alliance will be in existence as a legal entity
DataCite will be operating sustainably

4.5 Overall 2013-14 Outcomes

ANDS has been operating since January 2009. In that time ANDS has built a consensus on the importance of research data and research data infrastructure. In 2010-11, ANDS created a number of national research data services and engaged with a large number of organisations to start the realisation of the Australian Research Data Commons. In 2011-12, ANDS expanded and deepened that institutional engagement, as well as added to the range of available services. In 2012-13 ANDS consolidated its national services, introducing data citation services; continued to partner to build infrastructure through funded projects and selected engagements; strengthened the communities of practice and was a significant leader in policy development nationally and internationally.

In 2013-14, ANDS will build on this infrastructure that has been created. By July 2014, ANDS will have in place additional and enhanced national services:

- A vocabulary service
- Citation identifier service
- Integrated ANDS systems with ORCID researcher identification service
- Research Data Australia used as a resource by other eResearch infrastructure partners
- Functionality added to Research Data Australia to support national collections
- 8 new training programs on ANDS national services
- Documentation and training material to support uptake of 9 areas of importance to data management and curation
- Improved linkages to publications, services and grants and activities to further improve the context in which data can be discovered.

By July 2014 there will be further support for institutional research data infrastructure in place:

- 20 programs of work to extend infrastructure to support institutional research data assets
- Support showcasing of institutional data assets through the ability to highlight collection strengths
- Coordinated 5 community events to support these outcomes.

By July 2014 there will be an established environment of national collections that will make available:

- 12 collections of national significance from individual institutions
- 4 distributed national collections on themes of research importance in Australia
- Established partnerships with RDSI, the nodes and service providers to coordinate access to data and associated services and tools regardless of location
9 topic pages in RDA in areas of demand by research

ANDS will continue to take a leadership role in international data activities. It is a founding partner in Research Data Alliance and will be heavily participating in the second plenary and hosting the third in the coming year. It will also be providing support for the activities of the working groups associated with Research Data Alliance. ANDS will continue its active engagement in the DataCite consortium and further its relationship with ODIN. Additionally ANDS will be coordinating its efforts with its international counterparts specifically in the UK and Netherlands.

These outcomes, in concert with previous year’s activities, will produce the following desirable consequences.

Better Data

- A national resource of managed, connected, findable and re-useable research data now exists that previously did not.
- More data are being automatically captured from a wide range of instruments, ranging from small sensors to major national facilities.
- Data providers are increasingly seeing the value in providing their data through richer shared environments.
- Data are now visible to researchers that were never visible before.
- Researchers can more easily find and gain access to public sector data.

Better Research

- A cohort of leading Australian researchers has demonstrated the value of a richer data environment.
- Australian researchers can publish and cite their research data outputs.
- Larger questions and grander challenges can be addressed by a data commons that spans facilities, institutions, disciplines and sectors.
- New types of research are now possible that previously were not.
- Richer data environments can facilitate evidence-based policy.

Better Institutional Capacity

- Data can be captured, managed, and shared more easily at most research organisations in Australia.
- Research organisations and public sector agencies now provide policy support and invest more in data management.
Australian research institutions can more easily comply with the requirements of the Code for the Responsible Conduct of Research.

There is an established community of data management professionals.

Australian researchers and research organisations have better capability to operate and exploit the new research data infrastructure.

**Better Investment**

- There is more cost-effective reuse of research data software in Australia.
- Australia has improved ability to measure the quality and impact of all of its research outputs.
- Research Data Australia reduces duplication of effort and cost of unnecessary data acquisition by facilitating re-use.
- Common approaches, standards, and services have increased efficiency and coherence across the national data infrastructure investment.
- Established data management infrastructure now allows for stronger national policy and easier institutional compliance.

**Better International Engagement**

- Australia is seen to be a leader in research data infrastructure internationally.
- Australian researchers can more easily engage with international partners on data intensive research.

As a consequence of these individual outcomes there is an over-arching outcome; Australian researchers now have access to infrastructure that enable them to:

- systematically, reliably and authoritatively connect their research data to project, institutional and disciplinary descriptions; and
- simultaneously publish citable research data collections through institutional, disciplinary and national services.

This will ensure that Australia has a mature, globally leading capability in research data, making it a key locus for data intensive research. This capability will be demonstrated by leading researchers in a variety of disciplines to show the power of this infrastructure to enhance their research.

## 5 Program Engagement Strategies

ANDS succeeds by building strong partnerships that enable the changes it seeks. Generally its partners will have needs that require a response from a number of ANDS programs, and its partners should never need to navigate ANDS’ internal structures. Moreover the programs have strong
interdependencies. For a university to respond to the Australian Code for the Responsible Conduct of Research effectively there is a particularly strong need for efforts from the National Services and Institutional Engagement programs. Consequently from an external point of view partners should be engaged with ANDS as a whole, not a specific program within ANDS.

ANDS has developed a customer relationship approach that enables it to have a single point of contact for a given level in the organisation – this might mean that for a given university one of the directors or managers oversees the relationship, and one of the ANDS team is the point of contact for the university. In this way the university never needs to work out whom to talk to in order to discuss the challenges of data publication; rather the ANDS client liaison officer will ensure the appropriate conversations take place.

Engagements will also be quite varied in nature. ANDS is engaging with partners to directly change research data practice at universities, NCRIS capabilities, Publically Funded Research Organisations, government departments conducting research, and other locations of publicly funded research. ANDS is engaging with organisations that have a direct influence on the Australian research system – data providers and holders including government departments such as the ABS, NAA, GA, Cultural Collections organisations, policy and funding bodies, such as the ARC, NHMRC, The Committee of Australian University Librarians (CAUL), The Council of Australian University Directors of Information Technology (CAUDIT), AVCC, AGIMO, etc. ANDS sees NECTAR, RDSI, and NCI as important partners in delivering the vision of the Platforms for Collaboration. Finally and most importantly ANDS is engaged with government through DIICCSRTE.

ANDS’ forms of engagement will be equally varied – in some cases ANDS will have a staff member work alongside a staff member in the partner organisation so that together they can institute good data practises within that organisation in a nationally consistent manner. ANDS will do this for example with TERN and IMOS. In some instances ANDS will have several staff work intensively but for a short period of time with a partner such as it has done with the University of Newcastle. In some instances ANDS might simply build a feed to a data repository to capture collection information that is already locally held.

ANDS intends to support local engagement as much as is possible, consistent with the view that ANDS is seeking cultural change, not just technical change; therefore personal engagement and relationships are important. Staff have been appointed in most states working on the ANDS relationships based in that state. These state based staff are generally expected to be located at the state based eResearch organisations with local line management, and ANDS project management.

6 Promotion

ANDS is a complex, innovative research infrastructure program, the continued success of which requires an increased awareness of the function of ANDS amongst its partners and the recipients of its services. ANDS has reached an unprecedented level of activity and delivery of outcomes and thus the need to promote ANDS – its successes, its value and the impact on research – is paramount. ANDS will make ‘taxpayer impact’ a particular focus during this period in order to communicate the
return on the government’s investment to a wider audience. To provide a focus for promotion in this period ANDS has developed an external communication plan, which details the key messages, key audiences, communication channels and promotional strategies.

Another primary focus during the 2013-14 period will be to work with DIISCCRTE and other infrastructure partners on a co-ordinated communications approach to eResearch.

Key promotional activities and communication channels planned for the 2013-14 period include, but are not restricted to:

- **ANDS Website**
  - Continue to improve the restructured site to ensure it aligns with ANDS’ overall Communication goals, ands.org.au will be used to build communities through features such as a blog, webinars and links to communities of practice.
  - The website, through continuous content refinement will also act as a comprehensive, focused tool for ANDS’ partners and reinforce ANDS’ reliability and credibility.

- **Collaboration with infrastructure partners (National and State-based)**
  - Harnessing promotional activities such as road shows, workshops, media relations.

- **Media Relations**
  - Demonstrate the value of the Federal Government’s investment through taxpayer impact stories focusing on specific outcomes that have been achieved through the reuse of data which was made available through the research infrastructure ANDS developed.

- **Events**
  - Launches of major impact eResearch initiatives such as the Research Data Alliance and the Climate Change Adaptation Information Hub projects.
  - Collaborative events with other capabilities, in co-ordination with DIISCCRTE’s communication strategy.

- **Research Data Alliance**
  - Strong international engagement with a view to maintaining Australia’s position as a global leader in research data infrastructure.
  - Meetings, working groups, conferences, presentations.

- **Social media**
  - Twitter account (@andsdata). The ANDS Twitter account has been an effective communication channel for ANDS to communicate with researchers and the research community, both within Australia and internationally. Other social media platforms are under evaluation.

A key element of the ANDS external communication plan is ongoing communications with the community through various engagements. This approach utilises the expertise of four distinct areas
within ANDS to ensure that our key messages are being delivered to our key audiences through correct communication channels, utilising the extensive expertise within ANDS:

- Directors (including the Executive Director)—This group is the strategic ‘face’ of ANDS. They liaise with research institutions and organisations at a high level, both in Australia and Internationally.
- Outreach staff and the Capabilities Team—This group plays an operational role in communications. They communicate with practitioners and researchers within institutions; they provide advice, guidance and community building through training, knowledge sharing and community building activities.
- Communications Officer—based within the operations team, this role promotes ANDS to a broader audience, and specifically promotes the value of ANDS (i.e. the government’s investment) and the benefit to taxpayers. This role is also responsible for working with other infrastructure partners on the co-ordinated communications approach to eResearch mentioned above.

7 Access and Pricing

The mechanisms for deciding access and pricing will be consistent across the ANDS services. Generally speaking, ANDS will provide services for research purposes and aims to ensure the legitimate research use of those services will be free and access to the services open.

Software developed under the programs will be released as Open Source code, with the choice of licence and licensing conditions varying on a case-by-case basis. Documents produced or funded by ANDS will be made available as public documents, on a no warranty, royalty free basis using the CC-BY license. ANDS will maintain a register of software that is produced through its funded projects.

However content access and charging regimes belong in the hands of content providers, so that the access and pricing issue in ANDS relates to the rules under which content may be provided into the ARDC and therefore supported by ANDS utilities and other support activities.

ANDS services will be restricted to users who are non-commercial, and engaged in research, with the exception of Research Data Australia, which will be searchable by any interested party. Research Data Australia will not be used for the advertisement of paid services. Equally, the Identify my Data service will only be accessible for non-commercial use.

8 Governance, Management and Implementation

The Governance and Management arrangements for ANDS are described in the contracts for the NCRIS project and the EIF project, as well as in a separate Collaboration Agreement. These arrangements have been deliberately designed to ensure that the governance is as open as possible, consistent with the acceptance and management of risk by the lead agency. The Governance arrangements were established for the NCRIS contract, but DIISCCRTE, Monash University, the lead
agent, and the Steering Committee have agreed to use the same approach to governance and management for the EIF ARDC contract and other contracts as well.

8.1 Governance Framework

Monash University entered into an agreement with DIISCCRTE to implement the Projects, receive NCRIS and EIF Funds and be accountable to DIISCCRTE for execution and performance of both Projects. Monash University has established a Collaboration agreement with the Australian National University and CSIRO as partners in the projects.

Monash hosts and operates one of the ANDS Offices, which will be used to manage the Project. ANU hosts the other office that houses both ANU and CSIRO staff.

Monash appointed the independent Chair of the Steering Committee after consultation with DIISCCRTE and the ANDS partners and formally includes the independent Chair in the performance management arrangements of the Executive Director of ANDS. The Executive Director of ANDS is Dr Ross Wilkinson.

8.1.1 Steering Committee

The current ANDS Steering Committee comprises a minimum of four (4) and a maximum of eight (8) voting members, including;

(a) an independent chair appointed by Monash;
(b) one representative appointed by each of the ANDS Members; and
(c) such additional persons as the ANDS Steering Committee may agree, such as data provider, data policy and other specialist representatives.

DIISCCRTE has nominated a non-voting observer, Cheryl Kut or her nominee.

The processes of the ANDS Steering Committee will be as transparent as possible.

As at March 2013 the current ANDS Steering Committee Members are:

- Independent Chair: Dr Ron Sandland;
- Ms Cathrine Harboe-Ree (Monash University);
- Mr David Toll (CSIRO);
- Prof Robin Stanton (The Australian National University);
- Prof Mark Ragan (University of Queensland);
- Mr Paul Sherlock (University of South Australia);
- Dr Siu Ming Tam (Australian Bureau of Statistics);
- Prof Craig Johnson (University of Tasmania); and
- Executive Director (ex-officio): Dr Ross Wilkinson (Australian National Data Service).
It is anticipated that the ANDS Steering Committee membership can be expanded over time to incorporate any additional requirements of the Project.

8.1.2 Management structure

ANDS is currently managed by a full time executive staff comprising an Executive Director (located at Monash), and three Directors (a Monash Director, an ANU Director and a CSIRO Director) as currently agreed under the ANDS Collaboration Agreement.

Directors report to the Executive Director with regard to ANDS activities and to a nominated person in the host institution for administrative purposes (the Supervisor). The Supervisor is normally the host institution’s representative on the Steering Committee.

Directors normally have a high degree of autonomy within their areas of responsibility but work under the leadership of the Executive Director.

If there is disagreement or conflict between the Executive Director and a Director the matter is discussed with the Supervisor in the first instance, after which it can be escalated to the Chair of the Steering Committee and, if necessary, the Steering Committee.

Any alterations to this arrangement will be as a result of, and documented in, a revised Collaboration Agreement that takes account of this Project.

ANDS staff work collaboratively with each other and support activities across ANDS. Some will be located at ANDS Member institutions and others out ‘in the field’. These field locations may include members of the Australian Research Collaboration Services consortium, a Division of CSIRO or major data federating institutions.

ANDS staff within or appointed by an ANDS Member institution report to the relevant Director, or as otherwise negotiated for staff located in other institutions. These staff are appointed in consultation with the Executive Director.

If necessary, the Executive Director can direct, through the Directors, or other supervisory arrangements applicable at other institutions, the work of ANDS staff located in any institution.

The ANDS central office at Monash provides administrative support to ANDS and its staff, including communications, branding, and website maintenance.
8.2 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 key risks for ANDS in executing the Projects and the risk management strategies to be employed can be grouped into four major categories.

8.2.1 Political and Governance

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

Risk Factors:

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

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

Risk 2 – That the ANDS Project is not managed effectively

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

Risk Mitigations:
Management and planning processes have been put in place that include formal reporting and regular reviews to ensure the efficient conduct of the Project.

Regular meetings of Project staff are held to build a team approach. Communication structures in place to facilitate working together.

Staffing levels are monitored and adjusted as required.

Contracts and partnerships with state based organisations that host Project staff have been put in place that ensure that staff are clear about their role. Ensure that ANDS-funded staff based in organisations who are ANDS sub-contractors are not placed in a position of conflict of interest.

Ensure all projects commence by June 2012.

Ensure all late starting projects are closely managed.

Recruit replacement managers as soon as possible.

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

Risk Factors:

- University processes, focussed on student and supplier engagement, are not a good fit for “funding agency” activities. ANDS’ role as a “funding agency” in many of its programs has imposed additional requirements on the lead agent causing pressure on its staff to assist ANDS.

- ANDS EOI approach generates clusters of work with tight timelines that impact on specific university functions such as the Solicitors’ Office and Finance.

Risk Mitigations:

- Approval has been obtained for stream-lined approaches at Monash University to enable ANDS to work more effectively.

- Fund additional staff or specific work at Monash University to enable ANDS to work more effectively.

8.2.2 Relationships

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

Risk Factors:

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

- ARDC Applications program – Stakeholders might feel that the wrong decisions have been made in determining which projects to fund.

- Substantial new funding for other eResearch activities and reduction in new funding for ANDS projects.

**Risk Mitigations:**

- Maximise the effectiveness of connections between the Project and related PfC and other initiatives, including involvement of groups outside ANDS in the ANDS Policy Forum, the ANDS Technical Forum, and the ANDS Content Forum.

- Ensure that ANDS’ engagement with stakeholders meet their research data ambitions as well as ANDS’ requirements.

- Ensure ongoing, strong engagement with the Research Sector, including research infrastructure capabilities.

- All activity plans were developed after consultation with relevant stakeholders.

- Membership of the Steering Committee includes key stakeholders.

- Performance measurement for the Project should include effective stakeholder engagement.

- Effective communication of benefits to stakeholders.

- Provide a clear rationale behind the decision process for project funding.

- Communications activities have been increased to create awareness of the value of ANDS’ activities.

- ANDS effort has been increased in creating partnerships as compared to contracting.

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

**Risk Factors:**

- Partner produces outcomes of low quality or does not meet the requirements of the contract.

- Partner expends funds in a way that is not consistent with the EIF guidelines.

- Lack of effective arrangements in place to ensure the contracted services are provided to an agreed service level.

- Service providers see themselves as disconnected from the Project’s decision-making or strategic planning.

**Risk Mitigations:**
Formal procurement processes have been implemented to ensure that the requirements are understood and that potential suppliers meet the set criteria.

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

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

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

**Risk Factors:**

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

**Risk Mitigations:**

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

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

**Risk Factors:**

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

**Risk Mitigations:**

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

8.2.3 Impact

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

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

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

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

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

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

Risk Factors:

- ANDS is the only user and maintainer of actual or de facto standards, leading to inability to share maintenance and development costs.
- ANDS is the only source of development activity on particular technologies (RIF-CS, ORCA, ANDS Handle code).

Risk Mitigations:

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

8.2.4 Resourcing

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

Risk Factors:

- Limited availability of skilled staff (both within ANDS and in ANDS-funded projects) impacts ability to perform tasks funded by ANDS
- Funding uncertainty leads to potential for staff departures.

Risk Mitigations:

- Build a vision for the function of the ANDS for the longer term and communicate this to staff.
- Providing as much certainty to staff as is possible, and involving them in navigating the future.

8.3 Milestones for 2013-14

The main milestones for ANDS in 2013-14 are based on the activities and outcomes from the individual programs. Some of these will be derived directly from an individual program; many require activities across several programs to succeed. The milestones for 2013-14 are:
<table>
<thead>
<tr>
<th>Milestones</th>
<th>13Q3</th>
<th>13Q4</th>
<th>14Q1</th>
<th>14Q2</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Collections</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
<td>Commenced engagement with three institutions on establishment of collections of national significance with RDSI nodes or institutional infrastructure</td>
</tr>
<tr>
<td></td>
<td>Three topic pages and two themed distributed collections published</td>
<td>Services associated with themed distributed collections available via RDA</td>
<td>Collections published with initial three institutions</td>
<td>Continued engagement with next three institutions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Further two themes commenced for distributed collections</td>
<td>Services associated with themed distributed collections available via RDA</td>
<td>Collections published with next three institutions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Further three topic pages published</td>
<td>Further three topic pages published</td>
<td>Further three topic pages published</td>
</tr>
<tr>
<td>National Services</td>
<td>Training delivered for DOIs, ORCID, Metadata quality Materials developed: national collections, institutional engagements First ANDS post-projects national events held Functionality</td>
<td>Citation proof of concept projects: mid project reports received Community support webinars held: Geospatial, Data Citation, Identifiers, Vocabularies Briefings provided to DIICCSRTE, ARMS, Infrastructure Phase 2</td>
<td>Community support events held: National Collections Briefings and submissions provided to funding agencies on national data policy agenda Citation Identifier Infrastructure Phase 2</td>
<td>Citation proof of concepts: final reports received Research data management materials Phase 3 completed Move to community development of ANDS Software Products operational</td>
</tr>
<tr>
<td>Milestones</td>
<td>13Q3</td>
<td>13Q4</td>
<td>14Q1</td>
<td>14Q2</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td><strong>Milestones</strong></td>
<td>available in RDA to support participation in themed distributed collections</td>
<td>CAUL on national data policy agenda</td>
<td>implemented</td>
<td>Documentation of ANDS software suite complete</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Release of a refined set of services (Registration, Discovery and Data Collection Page Creation Infrastructure) based on User feedback</td>
<td></td>
<td>All ANDS post-projects national events held</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ORCID systems integration complete</td>
<td></td>
<td>Vocabulary service implemented</td>
</tr>
<tr>
<td><strong>Institutional Engagement</strong></td>
<td>Five institutional agreements on work programs for their research data assets</td>
<td>Further five institutional agreements on work programs for their research data assets</td>
<td>Further five institutional agreements on work programs for their research data assets</td>
<td>Further five institutional agreements on work programs for their research data assets</td>
</tr>
<tr>
<td></td>
<td>Successful conclusion of all Data Capture, Seeding the Commons, Metadata Stores and Applications projects (NOTE: most will have concluded by end of 13Q2)</td>
<td>Identification of some institutional assets</td>
<td>Identification of some institutional assets</td>
<td>Identification of some institutional assets</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Concluded agreed work on research data assets with initial five institutions</td>
<td>Concluded agreed work on research data assets with further five institutions</td>
</tr>
<tr>
<td><strong>International Collaboration</strong></td>
<td>Second Research Data Alliance Plenary takes place Linguistics</td>
<td>Another Working Group with significant Australian involvement approved by</td>
<td>ANDS hosts the Third Plenary in Europe Third Research Data Alliance</td>
<td>First Research Data Alliance Working Groups deliver data exchange solutions</td>
</tr>
</tbody>
</table>
### 8.4 Key Performance Indicators

A condition of the NCRIS Funding Agreement for ANDS is that “Key Performance Indicators (KPIs) acceptable to DIISCCRTE must be developed” (Attachment A, Section 5.1). The KPIs will enable ANDS to guide its behaviour, and DIISCCRTE and the steering committee to monitor the success of ANDS.

#### 8.4.1 Key Performance Indicator Series

1. The number and coverage of data repositories providing metadata feeds to the national registry compared to the number of data repositories.
2. The number and coverage of institutions and number of research groups with which ANDS has engaged
3. The number of institutions with research data management policies and practices consistent with ANDS recommendations
4. The number of times a search is initiated with an ANDS discovery service
5. The number of times an ANDS data page (defined below) is accessed
6. The satisfaction of researchers and partners (see below) with ANDS services as measured by an annual survey
7. The number of data access and sharing agreements with stakeholders – principally research institutions, government data agencies, government research agencies

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

8. The number of research data sets in harvestable repositories
9. The number of research data sets with persistent identifiers

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

These KPIs address ANDS objectives (refer 2.1) as follows:

- **The commons:** KPIs 1, 2, 4, 5, 7, 8, 9 and the long-term measure 10 address objective A.
- **Data management:** KPIs 3, 6, and ANDS’ long-term measure address objectives B and D.
- **Repositories:** KPIs 3, 8 and 9 address objective C.
- **Access:** KPIs 4, 5, 6, and 7 address objective E.
Use: KPIs 4, 5, 6, 7 and the long term aspirational measure 10 address objectives F and G. (Note – when KPIs 4 and 5 are being measured, not only use will be noted, but where it is initiated so that analysis can be done both within and across disciplinary use. The satisfaction survey will be qualitative, enabling an understanding of how well disciplinary, cross-disciplinary and multinational interaction is being facilitated.)

The form in which ANDS services are offered will be shaped by adherence to the guidance provided above. This guidance will be reflected in the business plans, and adherence to this guidance will be determined in discussion with stakeholders.

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, etc., and the research data using organisations, such as the Universities, the PFROs, and affiliates. Many organisations have both roles.

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

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

8.4.2 Key Performance Indicators for 2013-14

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 maintain at least 100 metadata feeds. This will cover at least 60 research data-holding institutions.
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 20 major Government data providers this year, and through them at least 50 research groups.
3. The number of institutions with research data management policies and practices consistent with ANDS recommendations: 31
4. The number of times a search is initiated with an ANDS discovery service: 50,000.
5. The number of times an ANDS data page (defined below) is accessed: 120,000 (now counting filtered page views using Google Analytics standard excluding robots spiders etc).
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.
7. The number of data access and sharing agreements with stakeholders – principally research institutions, government data agencies, government research agencies: we aim to maintain at least 50 agreements to make information available.

8. The number of research data sets in the ARDC: more than 60,000 collections

9. The number of research data sets with persistent identifiers: 14,000

10. The number of times a data set is reused and referenced – this will be tracked but cannot yet be reported.

These measures will provide indication of the effectiveness of the ANDS outputs this year and will be used as indicators to determine whether researchers can:

- systematically, reliably and authoritatively connect their research data to project, institutional and disciplinary descriptions, and
- simultaneously publish citable research data collections through institutional, disciplinary and national services.

8.4.3 Measuring our Performance against the Transformations

The KPIs just described provide a picture of ANDS’ overall performance, but do not explicitly show how well ANDS is addressing the key transformations that focus on our data improvement strategy. In this section we discuss ANDS performance in this light.

**Data that is unmanaged into managed collections:** ANDS has worked on the basis that data management is best centred at the institutions, so we have focused on improving research data management at institutions. Our KPIs 1, 2 and 3 provide an indication of this capability. More generally we see increase in the people, policies and procedures available at institutions to support research data management. We can also determine whether the relevant systems are in place to manage research data – does an institution have a metadata store, and is there software in place to support automated metadata capture? There has been a substantial improvement in institutional capability in this regard. Another important measure would be the number of funded research projects that have data management plans. This number is currently very small and is only likely to improve dramatically if there are changes in policy settings at funding agencies – this is where the work of the frameworks program is focused.

**Data collections that are connected:** There a number of ways for data collections to be connected: they may have a persistent identifier that enables them to be connected to publications, they may be described using standard terminology that connects the collections to similar collections, and they may have links to descriptions of the projects that generated them, parties associated with them, and services that are available to operate on them. They may be deliberately associated through an overarching collection, or through commentary. Our KPIs 7, 8 and 9 provide an indication of this capability, but the number of records describing collections in RDA – currently over 55,000 – provides an indication of a much more richly connected ARDC.

**Data collections that are discoverable:** Discovery of research data can occur in many ways: through an institutional portal, a disciplinary portal, through Research Data Australia, or through internet
search. KPIs 4 and 5 give an indication of discovery through Research Data Australia, but it is equally important that data is discovered through the form that makes most sense to a researcher. Whether water data is discovered through a Murray Darling Basin portal, through a TERN portal or through Research Data Australia is unimportant. That the data is discoverable is more important, and the number of feeds between different portals as indicated by KPI 1 is important to maximise discovery.

**Data collections that are re-usable:** There are a number of possible measures for this. One is the *level of reuse* of research data. Although the introduction of Digital Object Identifiers for data (such as the DataCite DOI, which ANDS provides) will make this easier over time, this is still quite difficult to do, and involves inherent lags. Another useful measure is the *enablement* of reuse. The Data Capture program funded the creation of 25 tools across various disciplines that support automatic capture of metadata to enable wider use. Data collections that have clear descriptions of licencing terms greatly facilitate integration and reuse. This has been another key focus of the Frameworks program, but statistics for the results of this focus have not yet been gathered. Finally, the behaviour of routinely reusing research data is greatly enhanced through demonstrations by leading researchers of what is possible. The number of compelling demonstrations of reuse is intended to increase substantially as a consequence of the Applications program over the period 2012-13. As an example, the Edgar system ([http://tropicaldatahub.org/goto/edgar](http://tropicaldatahub.org/goto/edgar)) enables researchers and the general public to access data on species occurrences held by the Atlas of Living Australia and visualise this in space and time (drawing on simulations of likely changes in habitat under conditions of climate change). Another 24 demonstrations of the value of bringing data together and reusing will be delivered by the Applications program late in the period 2012-13 and early in the period 2013-14.