



Australian National Data Service (ANDS)

BUSINESS PLAN 2011/12

INDEX

1	Executive Summary	3
2	ANDS Context and Approach	5
3	Status of Project	10
3.1	Establishment	10
3.2	NCRIS funding	10
3.3	Additional EIF Funding	10
3.4	Time Extension and Reporting	11
3.5	ANDS Programs and Relationship between NCRIS and EIF Activities	12
3.6	ANDS Principles	13
3.7	Scope	14
4	Research Infrastructure	16
4.1	Frameworks and Capabilities (NCRIS)	16
4.2	Seeding the Commons (NCRIS)	21
4.3	Data Capture (EIF)	28
4.4	Metadata Stores (EIF)	40
4.5	Public Sector Data (EIF)	45
4.6	ARDC Core Infrastructure (EIF)	51
4.7	ARDC Applications (EIF)	56
4.8	Overall 2011/12 Outcomes	60
5	Program Engagement Strategies	62
6	Promotion	63
7	Access and Pricing	64
8	Governance, Management and Implementation	65
8.1	Governance Framework	65
8.2	Risk Management	66
8.3	Milestones for 2011/2012	73
8.4	Key Performance Indicators	75
9	Appendix : NeAT Projects	79
9.1	NeAT Overview	79
9.2	ARCS-ANDS Agreed NeAT Governance:	79
9.3	NeAT Project List	80

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.

The driver for these initiatives is a desire for Australian researchers to have unprecedented access to data, enabling more efficient, and more new, research to be conducted in a richer data environment. To enable this, ANDS will focus on four data transformations – to managed, connected, findable and re-usable – that will enable Australia’s research data as a whole to become a nationally strategic resource.

In this document ANDS’ plans for the 2011/12 financial year are described in detail. This plan is in accordance with our extended strategic direction and builds on the work of commencing a very large number of activities in 2009/10. These are largely being conducted by others in partnership with ANDS to ensure national reach, re-use opportunity and coherence. Rather than institutions just meeting ANDS’ goals, the team will encourage organisations’ research groups and researchers to realise their research data ambitions.

There are no substantial changes to the proposed plan of work as a result of extending ANDS to June 2013. However we prioritised project commencement in 2009/10 and as projects conclude in 2010/11 we will be doing two things: focussing on re-deployment of capture tools, and on the use of tools to exploit research data. We have had to refine the Metadata Stores program given the very wide range of activity taking place in institutions and the number of projects creating object level metadata.

Significant progress has been made to date and by June 2011 ANDS will have:

- Established several national services, a data collections registration service, a dataset identification service, a data collection description publication service, and Research Data Australia – a data collections discovery service
- Populated the ARDC with data collections that have been managed, connected, and have 4,000 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, pipes connecting institutional systems, and connecting to 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, and new tools that enable more effective re-use of research data

During 2011/12, ANDS will:

- Establish more national services: a data citation service, a gazetteer service, a researcher identification service, a research project identification service and an enhanced Research Data Australia – a data collections discovery service
- Populate the ARDC with data collections that have been managed, connected, and have over 10,000 collections descriptions discoverable through Research Data Australia, Google and other mechanisms
- Help embed coherent institutional research data infrastructure at most research institutions, including automated tools for capturing rich metadata from instruments, metadata stores, pipes connecting institutional systems, and connecting to national services, and connected appropriate discipline data stores and portals with RDA
- Demonstrate that researchers are publishing their data, citing their data and being cited, as well as demonstrating the effective re-use of data

By the end of 2011/12 it is expected that there will be many researchers who 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*. Most importantly ANDS will have engaged the research community to the extent that researchers see publishing their research data as their default 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 and Research (DIISR) has invested in improving the Australia’s research sector’s capability to use and re-use research data. This has been guided first by the NCRIS roadmaps and then by the document entitled *Towards the Australian Data Commons*¹ (TADC).

In support of this, *Towards the Australian Data Commons* identified a range of objectives for ANDS. These objectives are 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)

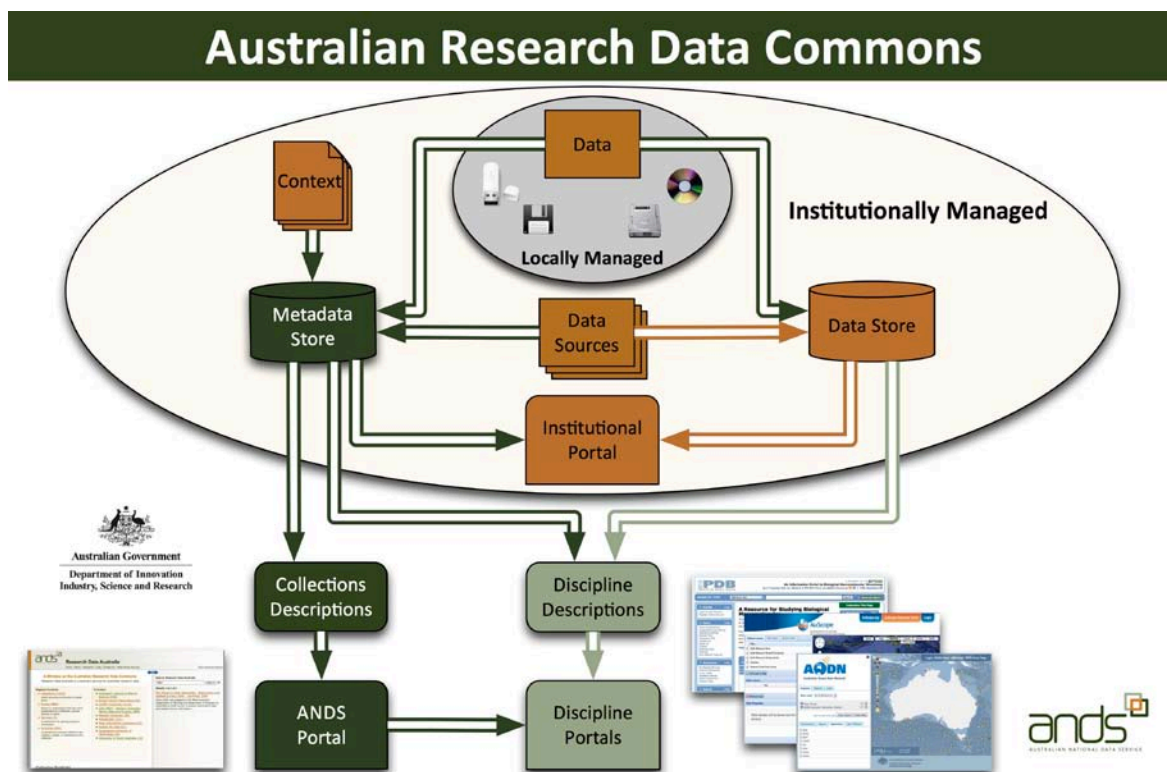
As a result of these goals, initial activity, and consultations, ANDS role is to enable Australia’s research data to be transformed:

From Data that are: <ul style="list-style-type: none">• Unmanaged• Disconnected• Invisible• Single use	To Structured Collections that are: <ul style="list-style-type: none">• Managed• Connected• Findable• Reusable
---	---

¹ Available online at <http://www.pfc.org.au/twiki/pub/Main/Data/TowardstheAustralianDataCommons.pdf>

This will form 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:



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 seven inter-related programs of activity (Frameworks and Capabilities, Seeding the Commons, Data Capture, Metadata Stores, Public Sector Data, ARDC Core and ARDC Applications). ANDS has also been funding specific development activity jointly with the Australian Research Collaboration Service (ARCS) under the banner of the National e-Research Architecture Taskforce (NeAT). These activities have formed part of a number of the ANDS programs, and are discussed in this document within the text describing those programs, as well as in Appendix A. The NeAT projects will all complete by the end of 2010/11, or very shortly thereafter.

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 DIISR for the NCRIS service and EIF project. This business plan provides a combined view of how ANDS intends to execute activities in 2011/12.

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 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, and
- 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 funded capabilities, especially expanding existing relationships with the Atlas of Living Australia, Terrestrial Ecosystem Research Network, the Integrated Marine Observation System, AuScope, AURIN, and the Australian Bio-security Intelligence 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, the outcome of ANDS activity is that for the first time ever, 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.**

This will ensure that Australia has a mature, globally leading capability in research data, making it a key locus for data intensive research.

It will be possible to do this as a result of ANDS and its partners developing a wide-ranging set of coherent software, policy and process outputs that support the Australian Research Data Commons.

Substantial infrastructure has already been created. By July 2011, ANDS will have in place the following national services:

- a data collections registration service,
- a data collection description publication service,
- Research Data Australia – a data collections discovery service, and
- a dataset identification service.

By July 2011 there will be coherent institutional research data infrastructure:

- tools will have been deployed to automatically capture rich metadata along with the data for a wide range of instruments,
- 4 institutions will operate metadata stores,
- 30 institutions will provide collections descriptions feeds to ANDS, both Research Institutions and Public Sector data holders,
- 4,000 collections will be available for discovery through Research Data Australia, and
- 2 discipline oriented portals will be cross connected to Research Data Australia.

By July 2011 the following tools, frameworks and capability will be in place to exploit the ARDC:

- improved institutional guidance for internal institutional data management,
- improved institutional guidance for responding to national instruments such as the *Australian Code for the Responsible Conduct of Research* will be available,
- institution wide research data management planning frameworks at 5 research institutions and 33 institutions with improved research data management,
- increased institutional capability for research data management with 100 more staff trained with research data management capability, and
- 10 new tools that enable more effective re-use of research data.

In 2011/12, ANDS will build on this infrastructure that has been created. By July 2012, ANDS will have in place additional and enhanced national services:

- A national gazetteer service
- A data citation service
- A researcher identification service
- A research project identification service
- An enhanced Research Data Australia

By July 2012 there will be further coherent institutional research data infrastructure:

- 60 tools will have been deployed to automatically capture rich metadata along with the data for a wide range of instruments
- 6 institutions will operate metadata stores
- 40 institutions will provide collections descriptions feeds to ANDS, both Research Institutions and Public Sector data holders
- At least 10,000 collections will be available for discovery through Research Data Australia
- 5 discipline oriented portals will be cross connected to Research Data Australia

By July 2012 the following tools, frameworks and capability will be in place to exploit the ARDC:

- Further institutional guidance for internal institutional data management
- Institution wide research data management planning frameworks at 10 research institutions and all institutions ANDS partners with have improved research data management

- Increased institutional capability for research data management with 150 more staff trained with research data management capability
- 20 new tools that enable more effective re-use of research data

We will continue to engage with the NCRIS Capabilities in a variety of ways – direct engagement on collections descriptions, common engagement with Government data providers, support on research data licencing, publishing research data collections held by the capabilities, especially by establishing collection descriptions feeds, providing another access point to the collections and capability portals.

3 Status of Project

The Australian National Data Service is one of the components of the Platforms for Collaboration (PFC) capability. Its status can best be understood in terms of its four sequential stages: establishment, initial NCRIS funding, additional EIF funding, and time extension.

3.1 Establishment

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. Details of these workshops and their outcomes are available at <http://www.pfc.org.au/bin/view/Main/Data>. 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). ANDS is part of the Platforms for Collaboration capability within the National Collaborative Research Infrastructure Strategy (NCRIS). 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 were identified.

3.3 Additional EIF Funding

In the May 2009 budget, the Commonwealth government announced a series of initiatives collectively labelled as Super Science³. 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

² <http://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf>

³ <http://minister.innovation.gov.au/Carr/Pages/SUPERSCIENCEINITIATIVE.aspx>

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 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 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 longer 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
- ANDS continued with existing partnerships - this means continuing to actively engage with the research institutions
- ANDS had the capacity to work with data champions
- ANDS maintained an ongoing capability of engaging with the sector

The project plan shows an un-even level of expenditure (as ANDS had already made substantial commitments) but does balance the need to engage with the sector over a longer period of time, and to demonstrate value early. This business plan describes the proposed activity over the second year of a three-year plan concluding in June 2013.

Another important variation to the ARDC project contract was agreed in March 2001 that replaced quarterly milestone reports being delivered individually with annual reporting that incorporates reports on ARDC progress as well as NCRIS progress.

3.5 ANDS Programs and Relationship between NCRIS and EIF Activities

There are now seven programs that have been established to meet the aims of ANDS and to create the infrastructure needed for the ARDC. The total set of programs therefore comprise:

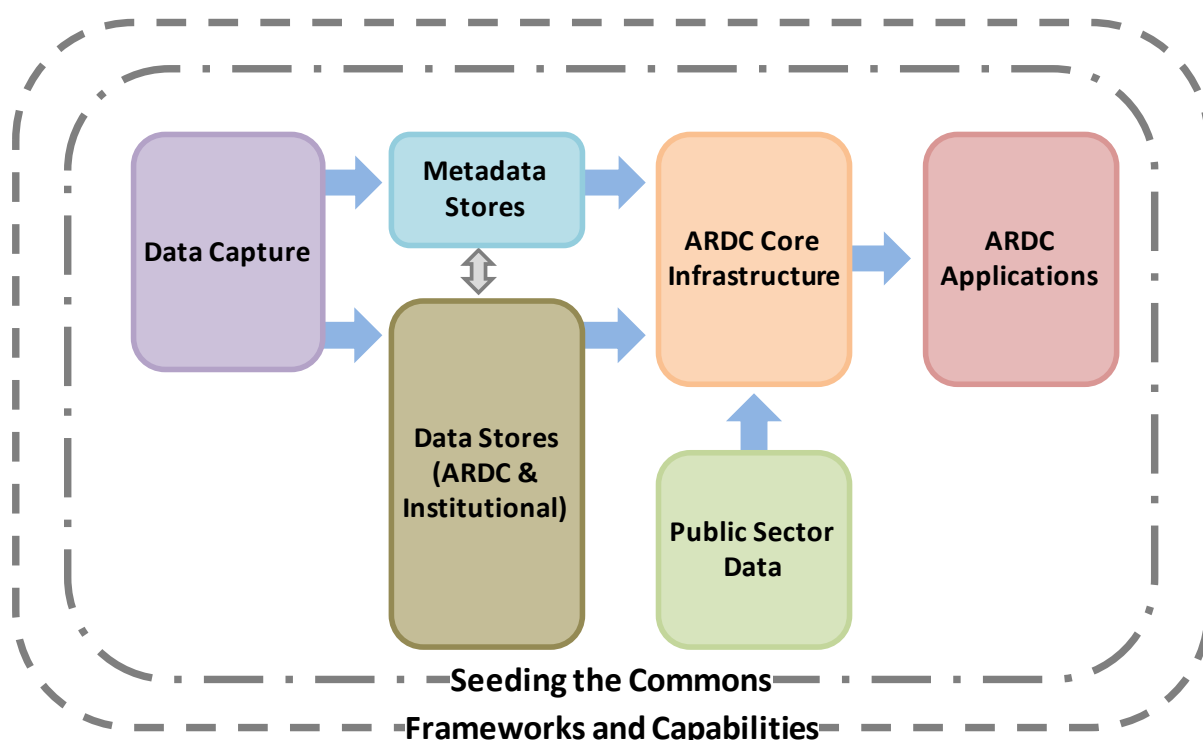
- **Frameworks and Capabilities** – an ANDS driven program ensuring that institutions have the capability and the research system have the structures in place to enable researchers to manage, publish, share and re-use research data (NCRIS funded)
- **Seeding the Commons** – an institutionally based program to ensure that well managed data is made available through the ARDC with a focus on data that cannot be automatically captured (NCRIS funded)
- **Data Capture** – an institutionally based program to automate the capture of data and metadata from instruments in data intensive research (EIF Funded)
- **Public Sector Data** – an outsourced program of making more public data collections visible and available through the ARDC (EIF Funded)
- **Metadata Stores** – an institutionally based program that enables metadata to be stored coherently across an institution that supports data management, publishing, sharing and re-use (EIF Funded)
- **ARDC Core Infrastructure**– an ANDS driven program that puts in place the national services that enable research data to be published and discovered (EIF Funded)
- **ARDC Applications** – a program that develops tools and services to support demonstrations of the value of exploiting data in the ARDC (EIF Funded)

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

Programs	EIF (\$M)	NCRIS (\$M)	%	Focus
Frameworks and Capabilities		4.40	6.0%	ANDS
Metadata Stores	6.54		9.0%	Institutions
Data Capture	18.47		25.4%	Institutions
Seeding the Commons		13.10	18.0%	Institutions

Public Sector Data	6.85	9.4%	Contractors
ARDC Core Infrastructure	7.8	10.7%	ANDS
ARDC Applications	9.87	13.5%	Contractors
EIF Project Management	1.83	2.5%	ANDS
Project Office		4.00	5.5% ANDS
Total (\$72.85M)	51.35	21.50	100%

The next diagram shows how the NCRIS programs complement and inter-relate to the creation of the ARDC Core.



3.6 ANDS Principles

In responding to the new objectives and program requirements, ANDS continues to follow these 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 directly 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 Australian Research Collaboration Service (ARCS), the National Computational Facility (NCI), and the establishing activities in National eResearch Collaboration Tools and Resources (NeCTAR) and 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 this 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.

3.7 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, and
- members of the cultural collections sector (galleries, libraries, archives and museums).

As a component of Platforms for Collaboration, 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 will need to be 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 and CAUL. 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.

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, and
- training and training materials that enable the infrastructure to be used well.

To deliver that infrastructure a combination of national services and coherent institutional services is needed. ANDS is creating 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 are also able to 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 seven programs to deliver on its NCRIS and EIF commitments:

- Frameworks and Capabilities
- Data Capture
- Seeding the Commons
- Public Sector Data
- Metadata Stores
- ARDC Core Infrastructure
- ARDC Applications

These activities 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 Frameworks and Capabilities (NCRIS)

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

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

program as a whole is the desire for research organisations and research groups to have effective policies around the full lifecycle of data management.

4.1.1 Program Aims

The Frameworks activities within the program aim to support new approaches to data-intensive research by strengthening the overall policy context for, and facilitating the emergence of, the ARDC. The Capabilities activities aim to improve the level of capability for data management, data-intensive research and associated technologies across Australia by partnering with willing institutions and NCRIS facilities to improve core data competencies. This program aims to directly support the ANDS IT services provided through ARDC Core as well as the institutional infrastructure projects managed through the Access to Public Sector Data, Data Capture, and Seeding the Commons programs.

4.1.2 Program Overview

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

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

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

As cohesive networks of research data are increasingly regarded as an important and enduring part of the collaborative research infrastructure, the Capabilities activities will focus in particular on

building the capability of researchers and support staff to contribute to and better exploit national data infrastructure. The 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.

The Capabilities activities will lead to services such as consultancy, informal knowledge transfer, workshops, documentation, and training materials both directly and by re-enforcing local services. Staff from this program will work within an integrated engagement activity with staff from all other ANDS programs. ANDS will identify and engage the community of researchers and e-Research support services. These groups themselves are engaged in capability building, within their own institutions as well as their own staff.

4.1.3 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 have already been conducted,

Infrastructure Exploitation Support Area	Existing Support Infrastructure
Data Commons Re-Use Policy	ARC, NHMRC, DIISR preliminary agreements Membership of sector and government policy committees (CJCIOC, NCDS, SDMG, MCI)
Licensing Frameworks - implementation	Commonwealth GILF implementers group Membership of CJCIOC GILF oversight committee ANDS guides
Collection Description Publication	Content Providers' Guide ANDS Guides Training, events
Data citation	ANDS Guides
Institutional research data management infrastructure development	ANDS Guides Training, events
Data management plans	ANDS Guides Training, events
ANDS Infrastructure Usage Support	ANDS Guides Training, events

4.1.4 2011/12 Activities

The 2011/12 activities build on the work done to date, continuing the focus on data management capability, but with greater focus on data citation and licencing in this period.

Infrastructure Exploitation Support Area	Activities
Data Commons Re-Use Policy	<p>Research funding policy advocacy</p> <p>Membership of sector and government policy committees (CJCIOC, NCDS, SDMG, MCI)</p> <p>Publication of cost-benefit analysis and application of its results in advocacy, materials and workshops (e.g UA, DIISR)</p> <p>Public forum</p>
Licensing Frameworks - implementation	<p>Promotion of AusGOAL within NCRIS facilities and research organisations</p> <p>Coordination of the pan-ANDS licensing agenda</p> <p>Materials development, events, website development</p> <p>Liaison with CAIRRS copyright</p>
Ethics and data re-use policy	<p>Review of practice at Australian research organizations</p> <p>Public forum</p> <p>Materials development, events</p>
Data citation	<p>DataCite collaboration and contribution</p> <p>Proof of concept projects with mature archives and disciplines</p> <p>Documentation and events to support ANDS DOI services, events</p>
Data Re-Use Information	<p>Materials development to document and support metadata stores initiatives</p> <p>ABS Data Quality transposed for research sector</p>
Institutional research data management infrastructure development	<p>Contribution to pan-ANDS research data management infrastructure service offering for research organisations (policy, plans, frameworks...)</p>

Infrastructure Exploitation Support Area	Activities
Data management plans	Working with partners International liaison Monitoring Australian policy
ANDS Infrastructure Usage Support	Support for Australian researchers and research organisations building ARDC infrastructure or using ANDS National Services

4.1.5 Challenges

The key challenge for the Frameworks activities is how to work effectively through collaborators. Clearly the success of the ARDC depends on favourable adjustments to the policy settings throughout the data and innovation sector. In some cases new paradigms will emerge. 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. The challenge for ANDS is to become an effective voice for research data.

A related challenge for the Capabilities activities is to create and empower a community of data management practice. Research organisations and research service organisations 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 cannot act independently, so a major challenge for this program is to act effectively through that community.

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.

4.1.6 End of 2011/12 outcomes

By the end of 2011/12 ANDS will have:

- Contributed to institutional research data infrastructure by providing materials, events and support on a number areas of strategic importance:
 - Data citation
 - Data quality and information for reuse
 - Data Licensing
 - Institutional frameworks, policies and planning
 - Data management project plans
- Contributed to key NCRIS stakeholders capability to exploit their own infrastructure

- Progressed discussions with funding bodies on the guidelines for re-use of research data and contributed to a forum with them
- Promoted the use of harmonised licensing across the research and government sector
- Published a cost benefit analysis for research organisations and data collecting agencies and promoted its conclusions in the sector

4.1.7 End of 2011/12 Enduring Changes

By the end of the 2011/12 funding year ANDS will have contributed toward, or brought about, the following enduring changes:

- researchers will have improved access to the data of publicly funded data producing organisations
- ANDS stakeholders will have made significant progress on research data management plans
- research groups will have formed research data networks which have the potential to support an emerging ARDC
- research organisations and data collecting agencies will have begun addressing the issues of data citation and data re-use.
- research funding agencies will be supporting good data management and accessibility in their funded projects
- the enduring changes will be supported by sound and consistent policies
- research institutions and state-based e-research services have materials and processes to address the data related capability requirements of their staff and stakeholders
- there will be an initial group of trained staff to drive forward the research data commons
- a community of practice with a sense of common national purpose has emerged amongst key stakeholders of the ARDC
- a body of community accepted knowledge and best practice is in current usage
- fundamental issues of data management and usage are clearly recognised and addressed

4.2 Seeding the Commons (NCRIS)

4.2.1 Program Aims

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

4.2.2 Program Overview

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

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

These activities are focused on growing the commons, supporting partners, and providing national advice. This advice will complement the written guides and training provided in the Frameworks and Capability program. Specific contributions to growing the data commons will provide:

- advice on metadata standards and requirements to ensure that metadata is prepared in a manner consistent with ANDS' needs
- advice on ANDS service requirements to ensure that metadata is available in ARDC
- advice and support on requirements to enable the creation and sustainability of research data infrastructure
- funding of staff at partner projects and institutions to provide support for ANDS' goals
- identification of as much content as possible and making it discoverable

During 2010 ANDS devoted considerable resources to helping institutions define and describe work that can be undertaken in these areas. By the commencement of this plan, all these projects should be described and underway.

Those universities that did not receive funding as part of this program have been offered further support and advice by program staff.

Activities to work more closely with partners will provide:

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

ANDS will create a community of data managers through:

- continuing to support the several state based or related groups of data managers established during 2010-11, with effective communication channels between them
- training provided to Community members as required (in conjunction with Frameworks and Capabilities Program)
- expanding the ANDS Knowledge Bank, for use by the community
- capturing information about successes for dissemination within the community and beyond
- developing relationships with equivalent activities overseas to share approaches to data management systems that can inform ANDS

4.2.3 Infrastructure Created To Date

By July 2011, ANDS will have funded the development of the following infrastructure components.

During 2010/2011 projects funded under the Seeding the Commons program at Griffith, Flinders, QUT and UniSA were completed. All of these delivered collection records to the ARDC, as well as promoting the growth of data management policy within the institution. In addition to these projects, ANDS-funded work at Monash and Swinburne has produced advice and guidance material on data management policy and practice, which has been made available to the larger data manager community, through ANDS communication tools.

ANDS has been 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 (the ANDS Boot Camp) to promote the understanding of ANDS' expectations around data management. This work has been done in conjunction with the Frameworks and Capabilities program. ANDS has also been actively seeking to broaden the discussion through a series of state based ANDS Community Events – one of which was held in all the mainland states (except NT) during the first half of 2011. Feedback from these events will be used to refine the shape of future events to be held in 2011/12, for instance events focusing on software developer needs.

4.2.4 Agreed 2011/12 Activities

Through this program, ANDS will concentrate in 2011/12 on growing the data commons, working with ANDS funded projects, and creating a community of data managers.

To grow the data commons, ANDS will emphasise the recruitment of existing content into repositories, identifying existing repositories of useful content, and making all that content discoverable through the ARDC. Where institutions with valuable existing content do not have the required systems, ANDS will work with them to improve their ability to store, describe, persistently identify and register their research data assets, in collaboration with the Metadata Stores program. If demand for this assistance exceeds available capacity, ANDS will develop a transparent process for allocation of ANDS resources. Where repositories (or federations) already exist, ANDS will assist with their integration into the ARDC. This may be work performed by staff in the program target (with ANDS advice if needed) or by ANDS staff (building on the technical consultancy expertise in the Frameworks and Capabilities program).

To enable ANDS to undertake this across the country, partnerships have been established with various eResearch support organisations. In these ANDS funds staff at the partners, whose role is to work with ANDS, the partners and other regional or discipline specific organisations to achieve the program's goals. It is ANDS' intention to co-fund these positions in 2011/12 with selected partners. Support for part of the program will also come from within the ANDS team. This includes both work undertaken through partners, and provision of specialist advice.

The availability of EIF funding, and the consequent reorganisation of ANDS has freed up considerably greater resources for use in the Seeding the Commons program. However, the amount of funding available for ANDS is still insufficient to provide a data management solution across the entire

research-producing sector. To focus funding effectively, a large number of Expressions of Interest were invited to understand better the needs and aspirations of institutions in this area. This provided funding for institutions to better understand their data management needs, their current data holdings, and put in place systems to manage and exploit this data. This can then be fed to the ARDC. ANDS staff will continue to work with these projects to ensure that the work is consistent with ANDS needs and recommendations. Any systems that are implemented will need to manage data and metadata that is captured, stored, persistently identified and made accessible (initially at collection level) through the ARDC. These institutional engagements will be undertaken with an awareness of the tensions between international disciplinary practice and national or institutional mandates.

ANDS has either entered into contracts (or has substantially agreed on project descriptions) for Seeding the Commons projects at the following institutions. The projects have generally taken one of three forms:

- **Broad:** a wide audit of available data, and of policies currently in place, with a view to describing data and creating wider policy
- **Exemplar:** working with exemplar data collections within the institution, with a view to applying the lessons learnt to other areas, and to the broader data management policy framework within the institution
- **Combined:** working with the Data Capture projects funded by ANDS, to apply the lessons learnt and help create a broader data management policy framework

Institution	Project Description/Name	Form of Project
ANSTO	Scientific Information Architecture	Broad
Australian National University	The ANU Seeding the Commons Project	Broad
Central Queensland University	CEM Core Data Curation project	Exemplar
CSIRO	Seeding the Commons: Enabling CSIRO's Biological Collections for the ARDC	Exemplar
Curtin University of Technology	Researcher Centric Model for the management of Research Data	Broad
Deakin University	Description and discovery of research data collections available at Deakin University	Combined
Edith Cowan University	Data Management Plan and Policy	Broad
Flinders University	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.	Exemplar

	database	
Griffith University	Identifying and describing Griffith University's research datasets and making metadata available to the Griffith research community and the Australian Research Data Service	Broad
Griffith University/Macquarie University	National Linguistics Corpus	This is a cross institutional project with a disciplinary focus.
James Cook University	Systematic JCU Tropical data collection discovery and description	Exemplar
La Trobe University	Archaeological Database Development: The People and Place Project	Exemplar
Macquarie University	Macquarie University Seeding the Commons	Broad
Monash University	Monash University Seeding the Commons Project	Broad
Murdoch University	Integrating precision agriculture	Exemplar
Queensland University of Technology	Seeding the Commons Funding	Broad
RMIT University	Screen Media Research Archive	Exemplar
Swinburne University of Technology	Swinburne's "Watering the garden for the seeds to grow" project	Broad
The University of Melbourne	Seeding the Commons	Broad
The University of Queensland	Seeding the Commons Funding	Broad
University of Adelaide	Research Data Storage and Management	Broad
University of Canberra	Cross-communication and enhanced accessibility for research data management systems	Broad
University of New England	UNE's N.C.W. Beadle Herbarium Database	Exemplar
University of New South Wales	Research Data management Services	Broad
University of Newcastle	Newcastle Research Data Online (Seeding the Commons)	Broad

University of South Australia	Taking Australian Architectural and Built Environment Records into the Commons	Exemplar
University of Southern Queensland	Sustainable policy and procedure for capturing research data	Broad
University of Sydney	Seeding the Commons at the University of Sydney	Broad
University of Tasmania	Publication of collections into the ARDC by UTAS	Broad
University of Technology Sydney	Community Tools and Processes for Effective Data Management Planning	Exemplar
University of Western Australia	Seeding the Commons through research data management at the University of Western Australia	Broad
University of Western Sydney	UWS Seeding the Commons	Combined
University of Wollongong	Identifying and locating UOW data sets to seed the Australian Research Data Commons and the development of a supporting research data management policy	Broad
Victoria University	Research data framework	Broad

It is expected that program staff engaged in these institutional activities will continue to be closely involved with the institutional Data Capture projects.

ANDS will also aim to share lessons learned and examples of best practice across the sector to create a community of data managers. This will involve continuing to bring together data managers either directly employed by ANDS, or funded by ANDS, from across the country, to effectively share knowledge and experiences. To achieve this ANDS will need to continue to develop and update an effective set of internal resources to centralise the necessary information and then to disseminate it widely. It is expected that members of the community will both add to and learn from these resources. ANDS will also work with other programs, key researchers, local bodies and overseas institutions to identify tools and infrastructure that could be co-developed to improve the quantity and quality of the data that is managed, and increase the richness of the contextual information around the data that is available. Close coordination with the Frameworks and Capabilities program will also play a key role.

As the projects under this and other ANDS programs are completed it is expected that there will a wide range of services, documents and software made available. ANDS intends to examine these outputs with a view to adapting, redeploying or further developing them for use in other institutions or projects. ANDS staff will work with the institutions to achieve this. The process for undertaking this is described in the next section.

4.2.5 Process for Other 2011/12 Activities

Given the limited amount of remaining funding, decisions on future external funding will be limited to small scale projects, such as the capture of specific collections, targeted partners or the deployment of software and other outputs of the underway projects. This process will be further refined during the year. However, selection of these projects or on redeployment will be based on:

- Coverage of existing discipline areas (where those areas with less coverage will need to be targeted)
- The number of new collections that are likely to be created
- Availability of relevant resources within ANDS
- The ability of the potential partner to commit similar resources and to support the services developed or deployed without ANDS funding
- The development of a catalogue of ANDS tools, derived from work funded by ANDS and elsewhere, and their suitability at potential partner institutions

These decisions will be made within ANDS, based on the outputs of the projects listed above. As such, final decisions will be made in second half of 2011.

4.2.6 Highlights

The funding of projects at many institutions should allow the program to meet its goal of considerably increasing the size of the ARDC. It is expected that most of the contracts with ANDS under this program will complete during this period. This will provide data to the commons, and increase the data management processes and practices at these institutions. The wider range of ANDS programs funded under EIF should also be enhanced by the outreach role of staff in the program, with the Data Capture, and Public Sector Data programs expected to benefit substantially.

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

4.2.7 Challenges

The proposed work to be funded under the contracts mentioned above is often ambitious in scope, and there are a large number of projects involved. As such there is potential for these not being completed, or failing to meet all their goals. Assessing and monitoring the number of projects in a way that is both auditable and yet does not present a roadblock for the partners is an issue, although considerable work has been done to refine the processes for this.

Some of the material developed may not be suitable for redeployment, or may only be able to be redeployed with considerable extra expense.

4.2.8 End of 2011/12 Outcomes

By the end of 2011/12 ANDS will have:

- increased the amount of data discoverable and accessible through the data commons
- seeded the commons by integrating a number of strategic data sources and federations into ANDS registry and discovery infrastructure, thus increasing their visibility and accessibility
- worked with partners and stakeholders to meet their contractual requirements to ANDS, and deliver as per agreed timelines, and completed 32 projects
- established the underpinnings of an Australian community of data managers
- begun redeploying or assisting in the redeployment of software and other outputs developed within ANDS or drawn from the catalogue of tools.

4.2.9 End of 2011/12 Enduring Changes

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

- more accessible data is stored in more institutions and is discoverable through the ARDC
- particular target groups have significantly improved their data management practices
- the ARDC has much better coverage of a number of identified strategic data resources
- there is an set of solutions available for deployment/uptake at Australian research-producing institutions that meet the data retention requirements of the *Australian Code for the Responsible Conduct of Research*
- there is an established community of data managers, who will continue to act as a community without ANDS funding

4.3 Data Capture (EIF)

4.3.1 Program Aims

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

4.3.2 Program Overview

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

approach taken will be to partner with leading research groups and Super Science initiatives to augment or adapt data creation and capture systems.

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

4.3.3 Infrastructure Created To Date

The National eResearch Architecture Taskforce (NeAT) projects were conceived as part of the National Collaborative Research Infrastructure Strategy (NCRIS) under Platforms for Collaboration (NCRIS 5.16). They were designed as a way of creating infrastructure that responded directly to the needs of particular discipline communities. The following projects were funded under the Data Capture program and are expected to be complete by or shortly after July 2011:

Project	Project Description
DTS: Data Transfer Service	This project worked with ARCS Data Services team to develop a general-purpose data transfer service with initial deployments at NCRIS Characterisation facilities.
Human Variome: Software and Data Support for the Australian Node of the Human Variome Project	The Human Variome project is creating a national data repository called the Australian Human Variome Database (AHVD). The database will hold and provide access to information on genetic variations associated with human disease that have been characterised by Australian laboratories and clinics. The project will develop services to enable submission of laboratory and clinic data to the AHVD using existing organisational workflows.
BioFlows: Bioinformatics Workflows	The Bioflows project is providing a simple Web-based workflow tool that enables life sciences researchers to specify genomics and proteomics workflows that can be executed on the ARCS Compute Cloud and interface with the ARCS Data Fabric. The system is deployable as an “appliance,” with required software, middleware and server hardware able to be installed at a site and managed remotely, if required. The appliance can interface with local high performance computing systems and/or submit compute jobs to the ARCS Compute Cloud. The appliance concept is being tested with trial deployments at the Bioinformatics Facility at Murdoch University, the Queensland Facility for Advanced Bioinformatics and the Life Sciences Computation Centre in Victoria.

Aus-e-Stage: Collective Intelligence and Collaborative Visualisation for Creative eResearch	The Aus-e-Stage project is developing two new visually interactive services for exploring information in the AusStage database. It is also creating the capability to generate a new data set of immediate, on-location responses from spectators of Australian performing arts.
Biosecurity: Biosecurity Collaboration Platform	This project has implemented a collaboration platform at the CSIRO's Australia Animal Health Laboratory (AAHL) facility that comprises two nodes – one on each side of the bio-containment barrier. This will greatly assist in the flow of complex information across the containment barrier from a variety of data sources including pathology and microscopy systems, live in-vivo animal experimental data (e.g. heart rates) and data from simulation models and historical information in both visual and written form. This platform is expected to have broader applicability within the NCRIS Australian Biosecurity Information Network
PODD: Phenomics Ontology Driven Data Management	The Integrated Biological Sciences component of NCRIS contains two major Phenomics initiatives: the Australian Plant Phenomics Facility and the Australian Phenomics Network. These facilities have common requirements to gather and annotate data from both high and low throughput phenotyping devices. The PODD project is delivering a data management service that can handle multiple phenotyping platforms and data formats (text, image, video).
Remote CT: Remote Computed Tomography Reconstruction, Simulation and Visualisation	The Remote CT project is developing a three-part service for 3D reconstruction and visualisation of Computed Tomography (CT) images. The service will be deployed at the Imaging and Medical Beamline at the Australian Synchrotron and the ANU micro-CT facility.
AusCover Workflow: Workflow Services to Enable a Large-Scale Temporal-Spatial Ecosystem Digital Information Service	AusCover is a component of the National Collaborative Research Infrastructure Strategy (NCRIS) Terrestrial Ecosystems Research Network capability. It focuses on organising remote sensing data sources and products for terrestrial ecosystems research. AusCover will enable – for the first time in Australia – the online storage of data sets in a form that makes them directly accessible to the user community. The AusCover Workflow project is providing easy-to-use workflow tools and services that enable researchers to process AusCover data sets using the ARCS Cloud Computing infrastructure. The same workflow tools also will allow AusCover data providers to more easily process raw satellite data to generate derived data products in the standard formats users require.

Other ANDS funded projects begun since the beginning of the ARDC project that have been or will be completed before the planning period:

Project	Institution	Project Description
---------	-------------	---------------------

ATNF Pulsar Data Management Project	CSIRO	The ANDS-CSIRO-ATNF Pulsar Data Management Project enables the discovery of, assessment of and access to Pulsar data observed at the Parkes Telescope.
Glycomics Repository	Macquarie University	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 APAF, with an OAI-PMH feed will be created to allow Research Data Australia (RDA) to harvest details of the collections stored within the GlycoSuiteDB repository.
Enhanced Metadata Capture for Sustainable Management, Sharing and Re-use of APN Histopathology Research Data	University of Melbourne	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.
An international antibiotic-resistance gene cassette database	University of NSW	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. The application makes this research accessible to many research teams that don't have the necessary computer skills to otherwise use it, thus increasing our impact on the scientific community.

Managing and Sharing Genomic Data	University of NSW	A new generation of DNA sequencers has recently been installed at UNSW, Southern Cross University and the Australian National University. These instruments can generate DNA sequence data 1000x faster than old technology, and can sequence the genomes of small organisms in a week. This project established databases for this DNA sequence information, so that users of the DNA sequencers can access their information in an efficient way. This will centralise DNA sequence data from these facilities, enable collaborative projects and facilitate data sharing. On publishing, research data and associated metadata can be made available for public use, contributing to the Australian Research Data Commons (ARDC).
Spatially Integrated Social Science	University of Queensland	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.
Aquatic Species Tracking Repository	University of Queensland	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, 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.
The Health-e-Reef Project	University of Queensland	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.

Development And Testing Of A Data Capture Tool For Social Datasets Being Used For Record Linkage	University of South Australia	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-automatic tool to facilitate capture of both data and metadata from the IWRI TRIFT V and Mastersizer 2000 instruments.
Clarke eHealth (Early Activity): Capture, management, re-use and discovery of breast cancer microscopy virtual images	University of Sydney	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.

These projects have not only delivered on their goals but have also helped ANDS to offer advice and refinements on how better to run other projects.

4.3.4 Agreed 2011/12 Activities

The Data Capture projects are designed to build infrastructure on at least one of three levels, although most work across these levels, as indicated in the table below. These levels are:

- **Instruments:** building infrastructure “pipes” between instruments and well supported data and metadata storage facilities.
- **Software:** to create these “pipes” and other infrastructure to enable better management and descriptions of research data and associated metadata
- **Management:** 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

ANDS has either entered into contracts (or has substantially agreed on project descriptions) for Data Capture projects at the following institutions. As noted above, some projects have been, or will be completed before the planning period. Shaded projects are still being defined at time of writing, but are expected to be agreed by the beginning of the planning period.

Institution	Project Description	Instruments	Software	Management
Australian National University	Earth Sciences	✓	✓	✓
	Optical Astronomy (Skymapper)	✓	✓	✓

	Phenomics	✓	✓	✓
	Humanities and allied disciplines	✓		✓
ANSTO	Metadata management for neutron beam instrument data - a joint project with Australian Synchrotron	✓	✓	✓
Australian Synchrotron	Meta Data Capture and Storage for the Three Mature Beamlines at the Australian Synchrotron - a joint project with ANSTO.	✓	✓	✓
CSIRO	A series of institutional data capture activities not yet determined but under discussion			
	ANDS-CSIRO-ATNF Pulsar Data Management Project	✓	✓	✓
Curtin University of Technology	A series of institutional data capture activities not yet determined but under discussion			
Deakin University	Filtration Membrane Fouling Data Collection for Water Treatment Research	✓	✓	
Flinders University	Automated measurement of the responses of wildlife populations to climate change		✓	✓
Griffith University	Smart Water	✓	✓	✓
	Adult Stem Cell & Neurobiological Microscopy Instrumentation and Research Data Management	✓	✓	✓
James Cook University	Tropical Data Hub	✓	✓	✓

Institution	Project Description	Instruments	Software	Management
La Trobe University	CMSS RLI Metadata Capture and Publication	✓	✓	✓
Macquarie University	Glycomics Repository	✓	✓	✓
	Papyri Data Capture			
Monash University	Research Data Management of the Monash Weather & Climate Program (Climate and Weather)	✓	✓	✓
	Biomedical Data Platform (Molecular Biology)	✓	✓	✓
	Tools for curating and publishing research data in the form of media collections (Multimedia Collections & ARROW)	✓	✓	✓
	Capture and publication of Australian ecosystem data from a network of measurement sites (Ecosystem Measurements)	✓	✓	✓
	Capture and publication of data on the history of adoption (History of Adoption)	✓	✓	✓
	Data Publication to Interferome (MIMR/Interferome)	✓	✓	✓
	Comprehensive Data Management for Microscopy Research Datasets	✓	✓	✓
National Imaging Facility	NIF Collections descriptions			
Queensland University of Technology	Greenhouse Gas Emissions from Australian Soils	✓	✓	✓
	Biodiversity	✓	✓	✓
	B150 Big Jam	✓	✓	✓
RMIT University	Data Capture from High Performance Computing Multi-User Environments	✓	✓	✓
University of Adelaide	Genomics Data Capture	✓	✓	✓
	Automated capture and publishing of data generated on high throughput plant phenomic platforms.	✓	✓	✓

Institution	Project Description	Instruments	Software	Management
University of Melbourne	Melbourne Neuropsychiatry Centre (MNC) Bioinformatics Development Project		✓	✓
	Youth Research Centre's Life Patterns Project: Longitudinal qualitative and quantitative survey data capture and reuse		✓	✓
	Video data in the Social Sciences. Optimising Metadata Capture, Data Sharing Procedures and Long-term Reuse		✓	✓
	Federated Neuroimaging Collections in the National Data Commons	✓	✓	✓
	Humanities and Social Science Research Data at the University of Melbourne		✓	
	Capture of Complex Data to Support Clinical Research in Cardiovascular and Neurological Medicine		✓	✓
	Founders and Survivors Project		✓	✓
	Enhanced Metadata Capture for Sustainable Management, Sharing and Re-use of APN Histopathology Research Data		✓	
University of New South Wales	Integrated Health Data Network		✓	✓
	ARDC Linked International Glycomics Repository & Instrument Data Capture			✓
	An international antibiotic-resistance gene cassette database			✓
	ANZNN Neonatal Data Capture Portal			
	Data capture and integration across multiple platforms			
	Managing and Sharing Genomic Data			✓
University of Newcastle	Data Capture for the Data Commons		✓	✓

Institution	Project Description	Instruments	Software	Management
University of Queensland	Spatially Integrated Social Science		✓	✓
	Microscopy/Microanalysis Image and Data Repository	✓	✓	✓
	DIMER Diffraction Image Repository	✓	✓	✓
	Aquatic Species Tracking Repository		✓	✓
	3D Anthropological and Archeological Collection Repository	✓	✓	✓
	The Health-e-Reef Project		✓	✓
	Linking the EMBL Australia EBI Mirror with the Australian Research Data Commons (\$1m)		✓	✓
University of South Australia	Development And Testing Of A Data Capture Tool For Social Datasets Being Used For Record Linkage	✓	✓	✓
University of Sydney	SKAMP Data Capture: astronomy	✓	✓	✓
	NSW TARDIS Node	✓	✓	✓
	AgDataCapt: Capturing Agricultural Data	✓	✓	✓
	AMMRF Live Cell Microscope Data Capture	✓	✓	✓
	Metadata Store/Aggregator	✓		✓
	FieldHelper: a workflow and tools for improving fieldwork data collection and submission to institutional repositories	✓	✓	✓
	Clarke eHealth (Early Activity): Capture, management, re-use and discovery of breast cancer microscopy virtual images		✓	✓
University of Tasmania	Redmap Australia	✓	✓	✓
	Data Capture of state-wide hydrological datasets	✓	✓	✓
University of Technology, Sydney	Maximising the Benefit from Data-Intensive Processes at UTS			

Institution	Project Description	Instruments	Software	Management
University of Western Australia	UWA1: Deployment and configuration of Institutional Metadata Repository		✓	✓
	UWA4: Integrated Data Capture for Characterization and Analysis	✓	✓	✓
	UWA2: Archaeological Rock Art Data Capture		✓	✓
	UWA3: Marine Ecology Video Capture and Storage		✓	✓
University of Western Sydney	Climate Change and Energy Research Facilities (CCERF)	✓	✓	✓
University of Wollongong	Biomechanics Data Capture Project System	✓	✓	✓
	Satellite data capture	✓	✓	✓

The projects all commence by providing a small number of hand-crafted feeds of Collections with associated Activities and Services. These feeds are used to inform the process of specifying the software that is required to automate the creation of this metadata, as well as to allow feedback on the metadata to be produced. Over the remainder of the project, the software will be specified, built, tested and documented. The expected runtime for most of the projects is 12 months from the date of contract signing, as such it is expect that most will complete within the current planning period.

4.3.5 Process for Other Activities

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

- Data capture infrastructure that is required to deliver data relevant to identified themes. This will require the Steering Committee to agree on the themes and accept recommendations from the Executive Director that particular infrastructure investments are required to deliver critical data.
- Data capture infrastructure that is required to meet the needs of one of ANDS' demonstrations of value or research champions. This will require the Steering Committee to agree on the instruments or researcher(s) selected and accept recommendations from the Executive Director that these particular infrastructure investments are required to support the success of this approach.

These allocations were originally expected to take place in 2010/11 – however delays with the EoI projects have pushed this back to 2011/12.

4.3.6 Highlights

Particular highlights of the Data Capture program so far have been the level of enthusiasm by many institutions to improve their data capture infrastructure, as demonstrated by the number of institutions that were ready to engage, the number of proposals submitted by some institutions, and the interest by a range of players in putting forward proposals.

A further highlight has been the completion of selected projects, as noted above. These projects have delivered infrastructure that supports the ANDS transformation goals, as well as offering the beginnings of the catalogue of tools that ANDS will further develop in 2011/12.

4.3.7 Challenges

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

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

Beyond all of this, the Data Capture projects have sometimes found it difficult to constrain the desires of the institutions (or particular researchers at the institutions) to do things that are somewhat data-related, but out of scope for this particular funding stream. It is clear that there is significant pent-up demand for help with data challenges.

4.3.8 End of 2011/12 Outcomes

The outcomes from the institutional activity funded under the Data Capture program will produce a greatly increased number of collections in the ARDC, as well as more data with associated rich metadata under management. These collections will span a wide disciplinary range and contribute towards the ANDS vision. Given the number of institutions being engaged and the number of projects within each institution, there is not sufficient space to list here all the anticipated outcomes in detail. In the case of the NeAT projects, the Data Capture infrastructure will contribute significantly to the amount of data made available for re-use both within the disciplines and across other disciplines, and will service as the beginning of the ANDS catalogue of tools.

The infrastructure resulting from both institutional and NeAT activity will also be available for adaption and redeployment in the remaining years of ANDS. This will take place under the Seeding the Commons program. It is expected that the majority of Data Capture projects will complete during this period.

4.3.9 End of 2010/11 Enduring Changes

The software infrastructure that will be developed and put in place within institutions will continue to make available data from the associated instruments until the instruments are replaced by models/manufacturers incompatible with the software.

The infrastructure and associated services being developed within particular disciplines under the NeAT projects will continue to contribute to the ARDC over time, as well as enable new kinds of research.

The thematic data concentrations will continue to deliver value to the Australian research community. Some of them may become obsolete as instruments increase their resolution, but much will be available to be built upon as research continues in these areas.

4.4 Metadata Stores (EIF)

4.4.1 Program Aims

The Metadata Stores program aims to assist institutions and disciplines to better manage the collection and object level metadata associated with research data outputs and associated entities.

4.4.2 Program Overview

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

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

ANDS is concerned with information about data collections and data objects, but importantly also with information about associated entities. These include parties (both people and organisations), activities (that produce the data) and services (associated with the capture of, and access to, the data). These associated entities serve as part of the rich context for the data collections, and also contribute to the information for discovery and information for determination of value. This rich

context will come from existing institutional systems via software infrastructure that might be thought of as pipes along which the contextual information flows. There also pipes between metadata stores and data stores, and between metadata stores and the ARDC Core infrastructure.

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

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

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

In addition to these, ANDS sees the need for an Object Store. This would hold metadata crucial to enable reuse of the data. It would meet the needs of an institution to manage object-level metadata for researchers whose needs are not met either by a Discipline Store or Instrument Store. This solution (and the pipes connecting it to data stores and the institutional Collection Store) is not currently part of the ANDS offering (nor deployed in any institutions) and will need to be adapted from existing software offerings. The final solution will need to manage metadata that is common to a range of disciplines, as well as specific metadata from particular disciplines.

As well as these different kinds of metadata stores, the data itself needs to be stored somewhere. This might be a local store (either just associated with an instrument or institutionally supported), one of the offerings that might be made available through RDSI, or an international discipline-focussed data store (such as the PDB or EMBL/EBI).

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

4.4.3 Infrastructure Created To Date

By the commencement of this Business Plan, ANDS will have funded the development of the following infrastructure components.

ANDS has funded three Combined Stores solutions from the Metadata Stores program. The CSIRO Data Management System has completed its initial development and now meets the needs of astronomy and water data. CSIRO is planning to extend it over the next year, progressively generalising it to the entire range of research areas within CSIRO. At this point it would fully meet the criteria for an institutional combined store. The SQUIRREL system is being built by Monash University to meet its institutional metadata store needs, managing both object and collection metadata. This development of this system is being co-ordinated with the co-ordinated metadata store activities at the Australian Synchrotron and ANSTO, and is being built on the same codebase. The SQUIRREL system should complete before the start of this Business Plan cycle – the Australian Synchrotron and ANSTO activities will still be continuing.

ANDS has funded two Collection Stores from the Metadata Stores program and both projects will have completed before the end of 2010/11. The VITRO solution is being developed by Melbourne University (based on the VITRO software and ontology from Cornell) and is being deployed by QUT and Griffith University as the basis for their Research Metadata Store Hub. The ReDBOX solution is being developed by USQ/ADFI and is being deployed by the University of Newcastle. Both solutions are being used to support Seeding the Commons funded projects at those universities.

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

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

Until the most recent Business Plan (2010/11), ANDS had been seeing a lack of demand for metadata stores solutions among some of our partners. The Data Capture and Seeding the Commons projects are building this demand as institutions see the benefits of better management for metadata about research data outputs.

4.4.4 Agreed 2011/12 Activities

All of the agreed 2010/11 activities that were originally funded through the Early Activity component of the EIF ARDC Project Plan are expected to complete. This includes the MACDDAP NeAT project. The exceptions that will run into 2011/12 are the co-ordinated activities at ANSTO and The Australian Synchrotron.

ANDS anticipates that a number of institutions will be reluctant to deploy two new pieces of software to manage collection-level and object-level information. Institutions may also prefer to have an integrated view of their research data outputs. ANDS therefore intends to fund the development of an expanded and generalised Combination Store, suitable for wider deployment, as well as selected deployment in a small number of reference institutions.

4.4.5 Process for Other 2011/12 Activities

As indicated in the Program Overview, ANDS will be spending in three areas:

- making available an Object Store solution
- adaptation/development of new classes of pipes
- deployment support for stores and pipes

For the Object Store solution, ANDS staff will first analyse existing solutions (both ANDS-developed and available from elsewhere) to identify the best fit for the required functionality. ANDS will then consider the most cost-effective and sustainable way to deliver the total functionality required. If any adaptation or extension is required, ANDS intends to approach development partners who have demonstrated their readiness, willingness and ability in past engagements rather than undertake an open call. This is because of the need to have this solution in place soon.

Installation of new pipes will be undertaken using a selection mechanism that starts with the EOI round ranking, but is also informed by the number of possible deployment candidates for any solution. Preference will be given to those solutions that will meet the needs of the largest number of the most research-intensive organisations. The same criteria for development partners will be used as for the adaptation/extension of the Object Store solution.

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

4.4.6 Highlights

There have been two highlights of the Metadata Stores program so far. The first is the way in which both the Seeding the Commons funded activities and Data Capture activities have generated interest in, and enthusiasm for, better management of information about institutional data collections. This appears likely to generate demand over this Business Plan period for deployment activity. The second highlight has been the way in which solutions that ANDS has funded have already been taken up by other institutions without ANDS-funded deployment support. This has been particularly noticeable for those institutions whose projects started later and who were thus able to take advantage of more mature metadata store offerings.

4.4.7 Challenges

The main challenge so far has been working out how best to meet the metadata needs of the Data Capture projects that are underway or about to commence when Metadata Stores projects had not yet commenced. The original timelines for the EIF funding of ANDS meant that we could not develop Metadata Stores first and then deploy them to support Data Capture activity. As a result, these two programs were commenced in parallel. This meant that most of the early Data Capture projects were not able to take advantage of well-advanced Metadata Stores solutions and had to create their own solutions. One approach that has proved successful has been to have one institution (NewCastle) act as deployment partners to USQ/ADFI as they developed an ANDS-funded solution (REDBOX), providing ADFI with real requirements and guiding the direction of the development.

4.4.8 End of 2011/12 Outcomes

By the end of the 2011/12 funding period, all institutions with which ANDS is engaging under the Data Capture and Seeding the Commons program should have been offered deployment support for an Object Store, and been engaged on their needs for pipe development/deployment.

The ANSTO and ASF coordinated activities will have completed and deployed production solutions to support the work of the identified instruments at those sites.

4.4.9 End of 2011/12 Enduring Changes

Every institution that wishes to describe the data collections that it holds will be able to do so by the end of the EIF funding period. A significant proportion of the work required to deliver on this will have completed by the end of 2011/12.

The metadata stores at institutions will continue to be available beyond the end of the initial ANDS funding period. These stores will continue to support institutions in managing the rich metadata associated with research data outputs and providing an institutional perspective across collections, activities and parties. The stores will also be candidates for feeding national registries (such as the NLA's People Australia or Research Data Australia) and for acting as a key component of institutional infrastructure.

4.5 Public Sector Data (EIF)

4.5.1 Program Aims

To develop the infrastructure necessary to ensure the establishment of automated feeds of rich collection level information from Federal, State and Territory government and non-government agencies into the ARDC and thus the Research Data Australia.

To engage with government agencies and provide them with funding or assistance to ensure that key data holdings are identified and the automation of data feeds into the ARDC are established.

4.5.2 Program Overview

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

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

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

4.5.3 Infrastructure created to date

ANDS has either entered into contracts, or has substantially agreed on project descriptions, for the following Public Sector Data projects:

Agency Institution Project	or or	Project Status and Description
---	------------------	---------------------------------------

Agency Institution Project	or or Project Status and Description
Aucope SISS Deployment	Deployment of the Spatial Services Information Stack (SISS) into 10 government and state or territory agency data providers including CSIRO, Geoscience Australia, BoM, ODSM and DPI(Vic) and to release 73 spatial information data collections into the Australian Research Data Commons. Negotiations are also underway to establish VerSI as a Victorian node to provide support to Victorian deployments of SISS. A Proof of Concept delivering archaeology data was successful and VerSI are keen to follow up during the 3rd quarter 2010.
CSIRO Water Resources Observation Network	Completed. 5 significant collections of data from the Murray Darling Basin Sustainable Yields projects have been exposed via Research Data Australia. Eight open source technology tools associated with the project were released for re-use, including: one to enable the ingest of data and metadata into the CSIRO Data Management System, a RIF-CS harvesting tool, a software library that provides programmatic access to the ANDS Persistent Identifier (PID) Service, NetCDF conversion tools, establishment of data storage and serving capability and a data and metadata management infrastructure. Additionally a web interface to provide data consumer access to data held within CSIRO was established and will be implemented for access to other data from the organisation.
AODN Data from National Research Vessels	This project has delivered 900 core data sets captured from research vessels (Southern Surveyor and Aurora Australia) published via the Australian Ocean Data Network portal and Research Data Australia (RDA) with further feeds of other data from the AODN portal under review. These data collections will expose the data of 6 commonwealth agencies with primary responsibility for marine data.
Australian Legal Information Institute (AustLII)	The original commitment has been met with over 400 databases and collections from courts tribunals and government agencies exposed via RDA. These data sets are principally the 'raw materials' of most legal research: legislation of all forms; Court and Tribunal decisions; treaties; official materials interpreting legislation; and reports proposing law reform. Exposure to AustLII databases provides the only free access, comprehensive national view of legal data of this nature. Additionally exposure has been given to LawCite, a completely automatically-generated case citator. An extension has been granted to December 2011 – to enable the gathering of a further 50 significant data collections.

Agency Institution Project	or or	Project Status and Description
Museums Metadata Exchange		This project has coordinated metadata from 18 museums across Australia including both major museums such as Powerhouse, state museums and the Australian Museum as well as smaller regional museums. The project will deliver a metadata exchange which will gather metadata from these museums and automate a feed into RDA. A test ingest of initial records into RDA test area to facilitate exposure of aggregated museum and history collections has already taken place. It is anticipated that on completion 700 collection records will be published. As well as the metadata store collecting metadata from other museums and the tools to automate the feed to RDA, the project will establish a vocabulary tool to support data standardisation in the museum sector.
National Archives of Australia		Proposal reviewed and value of commitment revised. NAA are re-assessing their intentions.
GeoSciences Australia		This engagement has enabled the automated exposure of data holdings from the GeoMet. ASDD and GeoCat data catalogues and will result in an initial release of approximately 700 data collections. It has enabled the identification of relevant data to be fed to other discipline portals, for example marine data. Automation of feeds has leveraged off the SISS deployment at GA.

The following NeAT project assigned to the Metadata Stores program will complete its final year of activity, as it has been funded through to the end of the 2010/11 funding period:

Institution	Project Description
National Criminal Justice Research Data Network	This project is developing minimum standards for storing and managing criminal justice data and providing for a single point of access to de-identified records sourced from criminal justice sectors agencies across Australian jurisdictions. NCJRDN has negotiated with ASSDA to host the data and the project will deliver 32 data collections from 11 agencies

4.5.4 Committed 2010/11 Activities

The following table represents engagements that are at the initiation stage of the process (a commitment has been made but the detail has not yet been finalised) for the Public Sector Data program of work:

Agency or Institution or Project	Project Status and Description
Australian Institute of Health and Welfare	The original proposal to automate the exposure of key holdings related to aggregated Australian health and morbidity data has been rescheduled at agency's request.
Australian Government Information Management Office	This is an ongoing strategic engagement to ensure alignment with GOV2.0 deployment taskforce
Australian Bureau of Statistics	Engagement activities have been rescheduled to accommodate staffing arrangements and will be commenced in Q3 2011
Bureau of Meteorology	The engagement activities reviewed in the light of NPEI activities as well as issues associated with varying demands for data and will commence in Q3 2011
Biogrid	This engagement commenced in Q2 2011 and will establish a feed to create relevant records and harvest into RDA. The data is real-time clinical, imaging and biospecimen data across jurisdictions, institutions and diseases from 14 major Australian hospitals. Potentially there are 20-30 collections covering thousands of datasets.
Bushfires CRC	This engagement has the potential for exposure of bushfire data from 48 Australian agencies federal, state and local government as well as educational institutions. It will involve the creation of website to enable registration of data, the delivery of collection descriptions followed by the infrastructure to provide access to the data. It is in the initiation stage

Agency or Institution or Project	Project Status and Description
Australian Antarctic Division	The intention of this engagement is to create an exemplar of collection presentation and transformation. It will provide data directly from the institution providing a vehicle for establishing tools to address the presentation of data from varying contexts. This work will build on the work done by the AAD to establish the Polar Information Commons.

Resources for the program will be increased over the 2010/11 timeframe to support the various engagement activities identified. The focus this year will be on providing government sector agencies with either the capability or the funding to ensure that the requirement to expose public sector data is achieved.

4.5.5 Process for Other 2011/12 Activities

To date the selection of government agencies with whom to engage has been as an opportunistic response to either identified researcher needs or as a follow through of enquiries initiated from the government sector itself. In the coming period a review of engagements will be mapped against an environmental scan. A survey will also be conducted to provide a high level view of data demand and this will be compared to the gap analysis and inform the prioritisation of approaches to further engagements. There will, however, continue to be a requirement for flexibility to accommodate opportunities as they may arise.

For capability-focussed public sector data access the engagement model will continue to concentrate on responding to the needs of those state, federal and territory government agencies where the organisation has indicated a desire for skills, knowledge and capability transfer. Implementation services will be established and offered by ANDS to offer assistance with the identification of key data holdings, and the implementation of automated data collection information into the ANDS ARDC by partnering consultants with government personnel for short periods of time with pre-arranged tangible outcomes. This will ensure that standard ANDS infrastructure requirements be applied to the data collections and that the agency has an opportunity to develop and build data management capability, standards and policies which would continue to make more public sector information available through ANDS well after the consultancy engagement has concluded.

For public sector institutions providing data feeds, the model of engagement will establish and fund contracted partnerships with representative agencies from federal, state and territory governments for the provision of agreed infrastructure components that will result in the release of rich collection level metadata to the ARDC. This work might be conducted by the agency or by partner such as AustLII for legal data, or AuScope for the Geological Surveys. It is expected that ANDS will fund partner engagements with those large government agencies that have existing capabilities and skills required to make their data holdings visible through the ARDC consistent with the selection process described above.

4.5.6 Highlights

Two of the contracted projects have been completed: CSIRO/WRON project and AODN. The remainder are established enough to provide learning for further engagements. One has been in relation to the requirement for data cleansing on the part of the agency to enable data description and metadata to be recorded and exposed. This has meant an additional effort as the requirements of the institution were sometimes not factored into delivery schedules. It is an important factor as it establishes the foundation for more efficient data management and exposure into the future.

Public Sector engagements have also tested some of the protocols around content provision as the business of these institutions differs from the university sector. While this has, on occasion delayed delivery, it has resulted in revised and strengthened processes that will enable more efficient delivery in subsequent engagements.

Another characteristic of Public Sector engagements is the “interconnectedness” of the data: data from a single source is often presented in several contexts. Example agencies in the current engagements are BoM and GA, with aggregators such as the NCRIS capabilities wanting to include some of their data in their discipline portals. The consequence for them is the requirement to implement their own data management in order to meet the demands effectively.

Projects agreed under the National eResearch Architecture Taskforce (NeAT) funding that were transferred to the Public Sector Data Program of work are expected to be completed by July 2011.

Agencies have welcomed an approach from the Public Sector Data Program to offer the services of skilled ANDS personnel for short durations to provide assistance with the identification and exposure of data holdings and build organisational data management capability. The onsite engagement with GeoScience Australia (GA) was the first of the ANDS engagements of this nature and acted as a use case for the development of a service package and for testing procedures such as defining collections and mapping metadata in order to build the tools to feed other aggregators such as IMOS, AODN and AuScope as well as RDA.

Projects that were endorsed and agreed under the National eResearch Architecture Taskforce (NeAT) funding, now transferred to the Public Sector Data Program of work, as well as the ‘early activity’ EIF funded projects are now complete.

4.5.7 Challenges

This is an ambitious program of work and engagement with the government sector is often difficult as not all agencies have internal policies and standards fully developed to support or enable the automated release of data. Many agencies are willing to engage but have no capability themselves to undertake such an exercise or lack the confidence to expose data holdings from within the organization. In the last 12 months this estimation has proved correct. Work done on the GA engagement has established some procedures and a checklist for facilitating this.

Most government agencies do recognize the need for good data sharing and reuse but are constrained by existing policies which determine the availability and release of existing public data. There is an overriding shift towards data management practices throughout the whole of

government at a policy level. This provides the opportunity for PSD to work in partnership with Capabilities and Frameworks to achieve the change necessary to effectively utilize the EIF funding.

ANDS has commenced negotiation with agencies with identified data holdings and has found the offer of skilled capabilities to be more attractive than the offer of funding. This has meant that the allocation of residual monies is difficult to determine until the negotiation of requirements has commenced, or is more fully underway, with public sector agencies and identified researchers.

4.5.8 End of 2011/12 Outcomes

By the end of the 2011/12 funding period ANDS will have:

- Commenced partnered engagements with at least 8 government agencies to establish infrastructure to expose data through the ARDC and RDA
- Made key collections from over 120 government data collecting agencies visible through the ARDC
- Made key cultural collections from at least 18 library, museums and other visible through the ARDC
- Made 480 public legal data collections visible
- Made 800 geological and other spatial data sets visible through the ARDC
- Supported key NCRIS capabilities in making public collections significant to them visible through both the ARDC and capability specific portals
- Strengthened partnerships with key NCRIS capabilities to provide a coordinated approach to the exposure of highly connected public data.

4.5.9 End of 2011/12 Enduring Changes

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

- Researchers will be able to more easily discover and access public sector data
- Public institutions will routinely make collections information visible through the ARDC
- The data management capability and awareness of the need to share public sector information with the research communities will have increased
- The government sector has increased avenues for the exchange of data holdings, thus increasing the capacity for stronger inter-agency analysis of data that could result in the production of more comprehensive advice to better inform public policy

4.6 ARDC Core Infrastructure (EIF)

4.6.1 Program Aims

- To ensure necessary technical and 24x7 operational services are established so that the content in repositories can be findable, re-usable and linkable, thus underpinning the development of the Australian Research Data Commons

- To ensure that services develop and evolve to meet changing data reuse requirements.
- To catalyse the emergence of core data commons infrastructure operated by government agencies and research organisations

4.6.2 Program Overview

The ARDC Core Infrastructure program provides fundamental services for a cohesive network of data collections and enables discovery, access and other value-add services across the resulting data commons. A technical consultancy is also available to assist integrating distributed data commons infrastructure at research and government instrumentality repositories with core ANDS utilities. The program has three areas of activity.

ANDS is establishing 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. This ANDS program also improves existing services and establishes new data commons utility services. Robust infrastructure will be established together with a service delivery framework that defines the roles and responsibilities of those providing, supporting and participating in the services.

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.

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

4.6.3 Infrastructure Created to Date

National Service	Status
National Data Collections Registration Infrastructure: Systems to allow information about research data collections to be registered and maintained dynamically through automated harvesting systems	In production
National Data Collection Discovery Infrastructure: The Research Data Australia portal that aims to provide a	In production

National Service	Status
comprehensive window onto the Australian Research Data Commons	
National Data Collection Page Creation Infrastructure: Allowing web pages to be created to advertise every registered collection	In production
Dataset Identifier Infrastructure: Allowing unique, persistent, and internet resolvable identifiers to be minted for objects in the research data commons	In production
Place Names Infrastructure: Web-service enabled gazetteer of Australian places established in partnership with OSDM/ GA	Under development
Researcher Identification Infrastructure: Infrastructure to enable access to definitive source information and identifiers for Australian researchers and research organisations (established with NLA) providing valuable information context to the ARDC.	Under Development
Research Activity Information Infrastructure: Web service enabled information system publishing definitive source information and identifiers for all funded Australian research projects (established with ARC and NHMRC) providing valuable information context to the ARDC.	In design
Standard Vocabularies Infrastructure: Infrastructure enabling the web-service publication, management and creation of standardised vocabularies used in research data laying a foundation for better linkages between datasets	In design
ARDC Infrastructure Establishment: Setting up best practice procedures for commissioning, establishment and management of national IT infrastructure commissioned through the ARDC project.	Ongoing

4.6.4 2011/12 Activities

National Service	Development Activities
National Data Collections Registration Infrastructure	<p>Multi-format support: Pre-harvest cross-walk</p> <p>Admin interfaces</p> <p>Architecture for third party deployment</p> <p>“Open source” management</p>
National Data Collection Discovery Infrastructure	<p>Provide further 'See Also' linking</p> <p>Allow further structured searching</p> <p>Scaling</p> <p>Architecture for third party deployment</p> <p>“Open source” management</p>
National Data Collection Description Infrastructure	<p>Update information and presentation design</p> <p>Scaling</p> <p>Architecture for third party deployment</p> <p>“Open source” management</p>
Dataset Identifier Infrastructure	<p>Integrate the ANDS DOI service with existing ANDS services</p> <p>Continue to contribute to the design and implementation of global data identifier and citation system</p>
Place Names Infrastructure	<p>ANDS is in partnership with responsible government agencies (GA/OSDM) to establish a robust national infrastructure that will allow place names to be validated by both individuals and software systems against an Australian Gazetteer (a directory that lists names of geographical place and features and includes spatial co-ordinates)</p> <p>Phase 2 (marine, boundary boxes, expanded dataset)</p> <p>Phase three planning (historical names, vernacular names, crowd sourcing...)</p> <p>Integrate Gazetteer services with existing ANDS services</p>

National Service	Development Activities
Researcher Identification Infrastructure	<p>Establish Infrastructure with National Library of Australia</p> <p>Project Benefit Realisation stage</p> <p>Integrate Researcher Identification Infrastructure with existing ANDS services</p>
Research Activity Information Infrastructure	<p>In collaboration with the Australian Research Council and the National Health and Medical Research Council create software components and services that will improve the accessibility and quality of information about research activities undertaken within Australia. When completed, the project will enable the public to discover and track the research undertaken through ARC and NH&MRC funding and to follow its connections to research outputs, both nationally and internationally.</p>
Standard Vocabularies Infrastructure	<p>Develop and implement software to enable the creation, management, and publication of human and machine readable 'terminologies' (also known as controlled vocabularies) for use by the Australian innovations sector.</p>
ARDC Infrastructure Establishment	<p>Provide technical assistance and design support to those organisations building the distributed infrastructure of the ARDC</p> <p>Provide specialist technical backup to members of the Australian research and higher education sector in the uptake of ANDS infrastructure</p> <p>Continue the establishment of ANDS infrastructure to support those services which have been implemented to date</p> <p>Continue to work to enable integration of ARDC Core Infrastructure with international infrastructure networks (such as NSF and JISC), including the registration of Australian data catalogues in Research Data Australia</p>

4.6.5 Challenges

For this program, within the 2011/12 period it will be most challenging to:

- instil confidence in the reliability and sustainability of these services,
- integrate fully with infrastructure of projects from other parts of ANDS

- completely ramp up national infrastructure
- encourage adoption of ANDS services

4.6.6 End of 2011/12 Outcomes

By the end of 2011/12 ANDS will have:

- implemented a DOI service; contributed to international infrastructure network
- designed, scoped and developed a terminology support service
- enabled additional universities and organisations to take advantage of ANDS utility infrastructure and assisted them with local implementations of data utility services
- established a community-source software stack for dataset collection description aggregation
- enhanced Research Data Australia to enable additional linkages, presentation, searching and geo-discovery features
- enhanced the ANDS Collections Registry to allow for a variety of data ingest mechanisms and information capture environments
- established the Researcher Identification Infrastructure with NLA
- established phase one of the Place Names Infrastructure with OSDM/GA
- established standard classification publication systems with ABS
- scoped further information infrastructure with other government agencies

4.6.7 End of 2011/12 Enduring Changes

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

- online and web services which support a collaborative Australian Research Data Commons will be available for use by the Australian research and higher education sectors
- distributed infrastructure of the data commons at research organisations and government agencies will be established and technically robust

ANDS will have formed strong technical partnership relationships with international research infrastructure networks.

4.7 ARDC Applications (EIF)

4.7.1 Program Aims

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

4.7.2 Program Overview

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

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

- result in data being transformed or integrated across multiple sources to produce new forms of information that enable innovative, high-quality research outcomes
- have broad rather than narrow applicability
- engage with national research capabilities.

These demonstrations of value will be in a range of sizes, depending on the opportunities that become available. Some of the demonstrations of value will consist of discrete activities that are aggregated together to meet the needs of an overall program of work. The sum of all the activity across this program will be designed to build a balanced portfolio of activity.

The Applications program is the last of the ANDS programs to commence (other than the funded NeAT activity). As such it is best placed to leverage the outputs from the other ANDS programs. Data Capture ensures that data and associated rich metadata is captured as close as possible to a range of instruments. Seeding the Commons is making a selection of existing data and associated metadata available from the bulk of Australia's research-producing universities to complement the new data coming from Data Capture. Metadata Stores ensures that this rich re-use metadata for data is managed and accessible to support the requirements of Applications. Not all of the data that might be needed for particular problems comes from research institutions, and so the work of Public Sector Data in making data and associated metadata available from government departments will be critical. ARDC Core provides the pieces of underpinning infrastructure to support discovery and citation of data sets for re-combination and analysis. Frameworks and Capabilities underpins Applications by putting in place the overall policy and practice frameworks to support better data management and re-use.

4.7.3 Infrastructure Created To Date

The Applications program has been funding a number of NeAT projects which have now successfully completed and delivered a wide range of useful tools to different disciplines.

The Annotation Services for Australian Literature (Aus-e-Lit) project has developed core data infrastructure required by the Australian literature portal Auslit to enable collaborative integration and annotation services for Australian literature. The Data Integration and Annotation Services in Biodiversity (DIAS-B) has developed core services required by the NCRIS Atlas of Living Australia

(ALA) capability to integrate data sources and use annotation to improve data quality. The Spatial Information Services Stack (SISS) project has developed spatial data infrastructure to meet the requirements of AuScope, the NCRIS capability supporting the geosciences research community (the SISS is the subject of a separate deployment activity funded by Public Sector Data). The ASSDA Services for e-Social Science (ASeSS) project has developed data curation, ingestion and analysis for social sciences data. See also the discussion of Enduring Changes below.

4.7.4 Agreed 2011/12 Activities

A set of climate change adaptation activities will have commenced by the start of 2011/12. The focus of these activities has been agreed in principle but individual projects have not been finalised. These activities are subject to the same criteria described in the next section.

4.7.5 Process for other 2011/12 Activities

As indicated earlier, ANDS plans to fund a balanced portfolio of activities under the Applications program. As a result, ANDS staff will seek to fund activities that over time contribute to this evolving portfolio. The nature of the portfolio and the kinds of activities that will be of interest will be the subject of a gradually expanding engagement between ANDS and institutions.

The criteria for elements of this expanding engagement can be described in terms of portfolio criteria and institutional criteria. These criteria are not mandatory, but ANDS intends to maximise as many of them as possible.

The criteria for the overall portfolio are that it contains activities that:

- will deliver early demonstrations of value
- align with the national research priorities
- are based on, and contribute to, research of recognised excellence
- address a range of NCRIS capabilities
- will lead to data being available in a form which was not previously possible, through transformation and/or integration across multiple sources
- exploit data that is already in place
- will enable the answering of new questions (previously not possible) and support innovative high-quality research outcomes.

The criteria for institutions are that they:

- have demonstrated the ability to deliver early demonstrations of value
- are already engaged with ANDS through a combination of both Data Capture and Seeding the Commons projects
- are producing excellent research in the required areas, as indicated by objective research measures
- have a demonstrated understanding of, and enthusiasm for, the project at researcher, management and leadership levels within the institution

The process for engagement will proceed as follows:

1. ANDS will over time use the criteria above to determine tranches of projects.
2. ANDS staff will meet with senior staff at potential candidate institutions to identify possible engagements and associated high-calibre researchers.
3. ANDS staff will then meet with the identified researchers to further refine the engagement and help document it as a fundable activity or set of activities.
4. The ANDS Executive Director will decide on activities that best meet the needs of the program.
5. ANDS will enter into a contract with the institution to fund the engagement.

The process will be spread out over time to ensure that ANDS is able to manage simultaneous engagements. The milestones for the Applications program (see section 9.3) describe the shape of this staggered engagement.

4.7.6 Highlights

All of the NeAT projects have completed and delivered services intended to meet the needs of their designated research communities, and are candidates for redeployment. The annotation service that has been developed in the Aus-e-Lit project has already been taken up by the DIAS-B project to meet the needs of biology researchers.

4.7.7 Challenges

The extension to the ANDS timeline required a re-think of the approach that the Applications program would follow. This re-think has taken some time, in part because of the need to accelerate the Data Capture program and ensure this major investment commenced as quickly as possible. The new Applications program is now more exciting, but also significantly more challenging. In particular, ANDS will need to carefully manage the expectations of the research institutions involved, both as to activity scope and timelines.

Inevitably with a limited budget and very large scope, another key challenge will be a defensible selection process that will be accepted by the research community, potential service creators, Monash University and DIISR. At the same time, selections need to be made rapidly to ensure that budgets are met, and the value of building on data is demonstrated in a timely fashion.

4.7.8 End of 2011/12 Outcomes

By the end of the 2011/12 funding period ANDS will have contracted eight separate activities, and completed four of them.

4.7.9 End of 2011/12 Enduring Changes

By the end of this funding period ANDS will have delivered a number of enduring changes to research communities assisted by the NeAT-funded process of delivering solutions:

- AusCover: users of satellite imagery (that could not do so before) will now be able to access more remote sensed data, simple workflows, and access to HPC
- Aus-e-Lit: Australian literature researchers are now able to conduct search across several data sources simultaneously (including full-text databases of Australian Literature) as well as annotate literary works with their own commentary
- DIAS-B: biology researchers will be able to better manage metadata about species and annotate information
- ASESS: social science researchers will have greatly improved access to archival data and analysis tools
- Biosecurity Collaboration Platform: bio-response researchers will be able to collaborate remotely across containment barriers and laboratories
- National Criminal Justice Research Data Network: a two-order of magnitude increase in researchers able to access data, and a substantial increase in available criminal justice data sets
- AMMRF Technique Finder: leading to increased visibility and use of AMMRF facilities
- Remote CT: has enabled a 100 times speed up of exploration of 2K x 2K mouse heart images; this has enabled explorations to be conducted by a single PhD student that previously required a team
- Phenomics Ontology Driven Database: richer metadata are now routinely and reliably captured and data can now be more easily made part of internationally significant collections because of the standards in use
- MACDDAP: more marine data is available to researchers and more researchers are using that data
- SISS: all States except NSW have deployed this at their Geological Surveys and the resulting data (together with rich workflows) is enabling the scale of questions to be increased

4.8 Overall 2011/12 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, the outcome of ANDS activity is that for the first time ever, 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.**

This will ensure that Australia has a mature, globally leading capability in research data, making it a key locus for data intensive research.

It will be possible to do this as a result of ANDS and its partners developing a wide-ranging set of coherent outputs that support the Australian Research Data Commons.

By July 2012, ANDS will have in place additional and enhanced national services:

- A national gazetteer service
- A data citation service
- A researcher identification service
- A research project identification service
- An enhanced Research Data Australia

By July 2012 additional coherent institutional research data infrastructure will be available:

- 60 tools will have been deployed to automatically capture rich metadata along with the data for a wide range of instruments
- 6 institutions will operate metadata stores
- 40 institutions will provide collections descriptions feeds to ANDS, both Research Institutions and Public Sector data holders
- At least 10,000 collections will be available for discovery through Research Data Australia
- 5 discipline oriented portals will be cross connected to Research Data Australia

By July 2012 the following tools, frameworks and capability will be in place to exploit the ARDC:

- Further institutional guidance for internal institutional data management
- Institution wide research data management planning frameworks at 10 research institutions and all institutions ANDS partners with have improved research data management
- Increased institutional capability for research data management with 150 more staff trained with research data management capability
- 20 new tools that enable more effective re-use of research data

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 both the Capabilities program and the Frameworks program. 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 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 who to talk to in order to discuss the challenges of data publication; rather the ANDS “relationship manager” 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 practise 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, CAUL, CAUDIT, AVCC, AGIMO, etc. ANDS sees AeRIC, 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 the DIISR.

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, so 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

A rich ARDC will become available if the management of research institutions support it, the information infrastructure that researchers rely on integrates with it, and researchers use it. Each relevant group needs targeting for promotion. ANDS continues to engage with institutional and government managers and administrators to ensure the existence of appropriate policy, training and support mechanisms. ANDS will work closely with information support organisations to create repositories and linkages that support the ARDC. ANDS will engage with researchers around the value of managed and discoverable data.

ANDS has developed a Communications Plan that provides for targeted communications (both ongoing and one-off) to the following audiences:

- Researchers – focusing on how to publish, manage and use research data
- Research Institutions focusing on policy and procedures and specific engagements
- Research infrastructure providers, focusing on engagements with ANDS, and mechanisms to connect to and exploit the ARDC
- Other partners such as NCRIS partners, Peak bodies, Government agencies, and international partners, which have individual communications

ANDS will use a variety of communication and promotion mechanisms. These will include the ANDS website (ands.org.au), email lists, wikis, a regular newsletter and through face-to-face communications such as conferences, seminars and community fora. Feedback arising from these communications, as well as ongoing consultation, will be used to inform future ANDS Business Plans.

In addition, ANDS will continue to engage with the following overseas institutions:

- The Joint Information Systems Committee (UK)
- The Digital Curation Centre (UK)
- SURFNet/SURFfoundation (Netherlands)
- Ministry of Research Science and Technology (New Zealand)
- The National Science Federation (US)
- NSF-funded DataNet projects (US)
- DataCite (Consortium)

ANDS will also be exhibiting and/or speaking at appropriate conferences in order to explain the ANDS approach and to develop wider partnerships.

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. 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 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 DIISR, 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 as well.

8.1 Governance Framework

Monash University has entered into an agreement with DIISR to implement the Projects, receive NCRIS and EIF Funds and be accountable to DIISR 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 DIISR 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.

DIISR has nominated a non-voting observer, Anne-Marie Lansdown or her nominee.

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

As at March 2010 the current ANDS Steering Committee Members are:

- Independent Chair: Dr Ron Sandland
- Ms Cathrine Harboe-Ree (Monash University)
- Mr David Toll (CSIRO)
- Mr Vic Elliott (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)
- Professor Craig Johnson (University of Tasmania)
- 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 four Directors (2 Monash directors, 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 will normally be the host institution's representative on the Steering Committee.

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

If there is disagreement or conflict between the Executive Director and a Director the matter should be 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 will 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 will report to the relevant Director, or as otherwise negotiated for staff located in other institutions. These staff should be 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 will provide 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

Risk Mitigations:

- Update the communications plans to ensure that the specific eResearch communities have input into specific projects and their outcomes before, during and after the projects are undertaken
- New diagnostic strategies need to be implemented and run to mitigate against failure
- Provide a central point where progress towards the ARDC can be tracked by metrics such as number of collections available, and numbers of datasets accessed. Clearly articulate the Project's message and brand.
- Engage carefully with communities to avoid perception (or reality) of not meeting its needs
- Ensure that the Project reflects the Government's expectations through constant dialogue
- Maintain close contact with key DIISR officers to ensure they provide input to decision making, including having an observer on the Steering Committee

Risk 2 – That the Project is not managed effectively

Risk Factors:

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

Risk Mitigations:

- Put in place management and planning processes that include formal reporting and regular reviews to ensure the efficient conduct of the Project

- Regular meetings of Project staff are held to build a team approach. Communication structures in place to facilitate working together.
- Negotiate with DIISR on appropriate administrative staffing for EIF funded programs.
- Staffing levels are monitored and adjusted as required
- Contracts and partnerships with state based organisations that host Project staff 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.

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

Risk Factors:

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

Risk Mitigations:

- Seek approval for stream-lined approaches at Monash University to enable ANDS to work more effectively
- Fund additional staff or specific work at Monash University to 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 – there is a risk of alienating clients/partners in determining which projects to fund

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 continuing wide consultation following the consultation on the Draft Final ARDC Project Plan
- Ensure ongoing, strong engagement with the Research Sector, including current and foreshadowed NCRIS capabilities
- All activity plans should be highly inclusive of 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

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:

- Implement formal procurement processes 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
- Put in place effective vendor and partner engagement approaches

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 causes a perception that ANDS initiated services will not continue

Risk Mitigations:

- ANDS seeks approval to expend existing funding over longer timelines (consistent with other Super Science funded activities)
- ANDS Steering Committee seeks additional funding for 2011/12, and 2012/13
- ANDS creates reliable sustainable services that are offered over the longer term by other long term service providers
- Strong contribution to DIISR Roadmap process will be a mitigating factor

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

Risk Factors:

- ANDS and Pfc partners' offerings are confused by possible users
- Relationship between ANDS and MARCS (such as Intersect) is not clear to users

Risk Mitigations:

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

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
- Researchers do not wish to share their research data
- Researchers do not trust the Project's data sharing and access control mechanisms
- Researchers are working with other collaborators who have confidentiality concerns over the data
- Existing data federations see insufficient value in making their data available

Risk Mitigations:

- ANDS will co-ordinate with RDSI and Institutional stores to mitigate this risk
- Recognise researchers through peer feedback for the deposit of data into the ARDC via increased citation – would need to be recorded and measured as a performance measure by the Project
- Effective communication of structures in place to ensure building of trust.
- Recommend that funding be linked to the provision of data via the ARDC as it becomes available
- Provide targeted assistance to data federations to assist with integration into the ARDC

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

Risk Factors:

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

Risk Mitigations:

- Ensure a nuanced and multi-faceted approach to exposing the Project's accessible data
- Work with ARCS and the Australian Access Federation to identify a simple set of standard access control policies
- Ensure that it is easy to re-purpose ARDC accessible data

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

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
- People with second order skills end up being employed because of staff shortages
- Limited tenure roles potentially on offer within the Project are not attractive to candidates
- Sustained high workload leads to staff burnout within the ANDS Management Team

Risk Mitigations:

- Commence recruitment early to mitigate delays in the commencement of activities
- Be highly selective in recruitment and favour quality of candidates over the quantity of candidates (do not fill jobs for the sake of it)
- Encourage secondment of staff at an institutional level
- Investigate non-traditional sources of potential staff
- Manage staff time and monitor levels of work

8.3 Milestones for 2011/2012

The main milestones for ANDS in 2011/2012 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 2011/2012 are:

Milestones	11Q3	11Q4	12Q1	12Q2
Frameworks and Capabilities	<ul style="list-style-type: none"> •Workshop held: data curation and preservation •Licencing Framework support activities underway •Cost Benefit Analysis published 	<ul style="list-style-type: none"> •Workshop held: Information Licencing •Materials developed: DOI, licencing •Citation proof of concept projects initiated 	<ul style="list-style-type: none"> •Ethics and Data Sharing review of practice under way •Forum held: Data re-use policies •NCRIS joint training activities confirmed 	<ul style="list-style-type: none"> •Materials developed: data re-use, data quality •Forum planned: Ethics and Data Sharing •Citation proof of concepts: first reports
Seeding the Commons	<ul style="list-style-type: none"> •All projects underway and project plans delivered to ANDS •Completion of 8 projects •Delivery of 800 records to Research Data Australia •8 new data management tools available via Open Source •3 Community events held •Knowledge Bank available to partners •8 policy documents available or in process 	<ul style="list-style-type: none"> •Completion of 8 projects •Other projects being monitored and assessed as required. •Delivery of 300 records to Research Data Australia •2 new data management tools available via Open Source •First version of catalogue of tools available •3 Community events held •7 policy documents available or in process 	<ul style="list-style-type: none"> •Completion of 7 projects •Other projects being monitored and assessed as required. •Delivery of 100 records to Research Data Australia •2 deployment engagements begun •3 Community events held •4 policy documents available or in process 	<ul style="list-style-type: none"> •Completion of 7 projects •Other projects being monitored and assessed as required. •Delivery of 250 records to Research Data Australia •3 new data management tools available via Open Source •3 Community events held •6 policy documents available or in process •2 deployment engagements begun

Data Capture	<ul style="list-style-type: none"> •All projects underway and project plans delivered to ANDS •Completion of 8 projects •Other projects being monitored and assessed as required. •Delivery of 70 records to Research Data Australia •new data management tools available via Open Source 	<ul style="list-style-type: none"> •Completion of 12 projects •Other projects being monitored and assessed as required. •Delivery of 800 records to Research Data Australia •15 new data management tools available via Open Source •First version of catalogue of tools available 	<ul style="list-style-type: none"> •Completion of 10 projects •Other projects being monitored and assessed as required. •Delivery of 300 records to Research Data Australia •new data management tools available via Open Source •Agreements reached on “Themes” •Feeds from previous projects are checked for activity 	<ul style="list-style-type: none"> •Completion of 12 projects •Other projects being monitored and assessed as required. •Delivery of 200 records to Research Data Australia •15 new data management tools available via Open Source
•Metadata Stores	<ul style="list-style-type: none"> •Work has commenced on adapting/developing an Object Store solution •Agreements reached with 6 institutions to redeploy an ANDS-funded Collection Store solutions 	<ul style="list-style-type: none"> •Testing commences for initial version of an Object Store solution •6 institutions have deployed an ANDS-funded Collection Store solution 	<ul style="list-style-type: none"> •An initial deployment of Object Store solution has taken place •Development/a daptation has commenced for a range of pipes 	<ul style="list-style-type: none"> •Further deployments of Object Store solution complete •Initial deployment of pipes (including testing) complete
Public Sector Data	<ul style="list-style-type: none"> •Environmental scan to determine demand for prioritisation of further engagements •Scheduled engagements with 4 agencies 	<ul style="list-style-type: none"> •Contracted projects to complete delivering 3000 data collections, 8 software items released into open source, 10 deployments of SISS, 1 metadata 	<ul style="list-style-type: none"> •Automated feeds of collection level data by initial 4 scheduled government agencies demonstrated as exemplars delivering 500 	<ul style="list-style-type: none"> •Automated feeds of data into the other discipline portals in ARDC has been provided by key government agencies •Automated feeds of

	<p>to commence</p> <ul style="list-style-type: none"> •Engagement with AODN to deliver 5000 data collections 	<p>store and exposure of data from 120 agencies (federal, state and local).</p> <ul style="list-style-type: none"> •Prioritised engagements commenced with 4 agencies 	<p>data collections</p> <ul style="list-style-type: none"> •Prioritised engagements commenced with a further 4 agencies 	<p>collection level data by 4 prioritised engagements demonstrated as exemplars delivering a further 500 data collections</p>
ARDC Core	<ul style="list-style-type: none"> •Citation Identifier Infrastructure phase one implemented •Researcher Identification Infrastructure complete 	<ul style="list-style-type: none"> •Registration, Discovery and Data Collection Page Creation Infrastructure phase two complete •Place Names Infrastructure complete 	<ul style="list-style-type: none"> •Vocabulary Infrastructure complete •ABS classifications complete •Citation Identifier Infrastructure phase two complete 	<ul style="list-style-type: none"> •Research Activity Infrastructure complete (ARC and NHMRC) •Data Collection Discovery and Page Publication phase three complete
ARDC Applications	<ul style="list-style-type: none"> •2 projects started •Discussions underway with a further 2 	<ul style="list-style-type: none"> •2 further projects started •Discussions underway with a further 2 	<ul style="list-style-type: none"> •2 further projects started •Discussions underway with a further 2 •At least one project completed demonstrating exploitation of the ARDC 	<ul style="list-style-type: none"> •2 further projects started •At least 3 projects completed demonstrating exploitation of the ARDC

8.4 Key Performance Indicators

A condition of the NCRIS Funding Agreement for ANDS is that “Key Performance Indicators (KPIs) acceptable to DIISR must be developed” (Attachment A, Section 5.1). The KPIs will enable ANDS to guide its behaviour, and DIISR 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 PFRAs, and affiliates. Many organisations have both roles.

Researchers have many partners in carrying out research and ANDS needs to satisfy these 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 2011/2012

1. The number and coverage of data repositories providing metadata feeds to the national registry compared to the number of data repositories. ANDS intends to build at least 80 automatic plus 100 manual metadata feeds. This will cover at least 35 out of the approximately 50 research data-holding institutions that we know about.
2. The number and coverage of institutions and number of research groups with which ANDS has engaged: ANDS will continue to engage with all Australian universities, PFRO's, and 4 major Government data providers this year, and through them at least 50 research groups.
3. The number of institutions with research data management policies and practices consistent with ANDS recommendations: 25
4. The number of times a search is initiated with an ANDS discovery service: Unknown – To date, it has been possible to browse RDA to find collections – this is the first year that will really require search to get to collections. As a result, this year we will monitor the use of search in order to have an appropriate KPI for next year.
5. The number of times an ANDS data page (defined below) is accessed: 300,000 in this year, up from the KPI of 100,000 for last year.
6. The satisfaction of researchers and partners (see below) with ANDS services as measured by an annual survey - no number can be given here, but a report will be provided.
7. The number of data access and sharing agreements with stakeholders – principally research institutions, government data agencies, government research agencies: we aim to strike at least 30 agreements to make information available.
8. The number of research data sets in the ARDC: more than 10,000 collections
9. The number of research data sets with persistent identifiers: 10,000
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 the overall outcome for 2011-12 is being delivered - the first time ever, 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.

9 Appendix : NeAT Projects

9.1 NeAT Overview

The National eResearch Architecture Taskforce (NeAT) projects were conceived as part of the National Collaborative Research Infrastructure Strategy (NCRIS) under Platforms for Collaboration (NCRIS 5.16). They were designed as a way of creating infrastructure that responded directly to the needs of particular discipline communities.

ANDS and ARCS have agreed, with the assent of DIISR, that the NeAT projects be funded 50% by ANDS and 50% by ARCS. DIISR have also approved the use of ANDS EIF ARDC funds for this purpose, as the aim of the NeAT projects is to build infrastructure. DIISR have indicated that they wish to continue with the governance arrangements detailed below, including payment by invoice against agreed effort.

The NeAT Projects are all in their final stages, and outcomes of these projects are described in the relevant Program descriptions. Their operation and financing were determined jointly with ARCS in accordance with DIISR requirements and the respective contractual obligations.

9.2 ARCS-ANDS Agreed NeAT Governance:

All NeAT projects should aim to establish services that are useful both for the discipline involved and as potential national services.

There should be only two levels of governance, where the distinction is clear between the governance and the deep technical and domain involvement needed for the project to succeed.

ARCS and ANDS have therefore discussed and jointly agreed on the management of NeAT Projects as follows:

- Each NeAT Project will have a NeAT Project Committee consisting of an ANDS representative (the Executive Director or delegate) and an ARCS representative (the Executive Director or delegate), representatives from any other institutions that would manage the enduring services provided by the NeAT Project, community nominated discipline representatives, a designated NeAT Project Manager (ex officio) and a prominent discipline leader as the NeAT Project Committee Chair. Where a suitable discipline Chair could not be found, the Chair will be either the ANDS or ARCS representative depending on whether the project was more ARCS or ANDS;
- Each NeAT Project will have a Project Manager selected by the relevant NeAT Project Committee;
- The Project Manager must be the person who manages the day to day work of the project;
- Project Managers must report to and be directed by the Project Committee;
- The governance structures of ANDS and ARCS will need to be satisfied with the Project Committee's management of the project in order to ensure the funds keep flowing, which provides the appropriate checks and balances and ensures accountability;

- At the start of the Project and subsequently once each quarter the Project Manager will attend a meeting chaired by the AeRIC Executive Director and attended by the Executive Directors of ANDS and ARCS and their nominees as well as the Project Managers of the other NeAT Projects;
- The Project Manager must meet no less than every four weeks with the Project Committee: in order to discuss the progress and evolution of the Project; to ensure that the Project is making optimal use of existing and planned services of project participants; and to ensure that the Project is being developed in a way consistent with the long-term delivery of the Services as per the project plan;
- Core responsibilities of each of the NeAT Project Committees include: overseeing and approving the design and implementation of an appropriate and relevant enduring service; and at the end of the Project identifying the key stakeholders and service providers to manage this enduring service into the future and to take over from the NeAT Project Committee.
- ARCS and ANDS will jointly review the progress of each NeAT Project every three months using their standard processes and the NeAT Project Committees would review their project every six months with a written report from the Project Manager. NeAT would review all NeAT Projects annually in September, beginning 2009, as part of the established NeAT processes.
- ANDS and ARCS will provide NeAT Project funds quarterly in arrears based on acceptable performance on a per EFT basis for each NeAT Project. The ANDS and ARCS quarterly reviews will be the trigger for either approving or withholding NeAT funding for that quarter from a NeAT Project or a component of that project as appropriate.

9.3 NeAT Project List

All NeAT projects are intended to be completed by the start of this business plan. Consequently any expenditure in 2011/12 will be final payment of late activity. The following table lists the NeAT projects that ANDS and ARCS jointly funded and the ANDS program that managed the project..

Project	ANDS Program
Aus-e-Lit	Applications
DIAS-B	Applications
MACDDAP	Metadata Stores
SISS	Applications
ASeSS	Applications
DataMINX	Data Capture
Data Transfer Service	Data Capture
AMMRF	Metadata Stores
Bio-Flow	Data Capture

Human Variome	Data Capture
BioSecurity	Data Capture
Aus-e-Stage	Data Capture
AusCover Workflow	Data Capture
NCJRDN	Public Sector Data
Remote Tomography	Data Capture
PODD	Data Capture
NeAT Project Office	Applications