



**FINAL AUSTRALIAN RESEARCH DATA COMMONS  
EDUCATION INVESTMENT FUND (EIF)  
PROJECT PLAN**

**FINAL AUSTRALIAN RESEARCH DATA COMMONS  
EDUCATION INVESTMENT FUND (EIF)  
PROJECT PLAN**

**CONTENTS**

<b>1. PROJECT OVERVIEW</b>	<b>1</b>
1.0 Background and Context	1
1.1 Definition	1
1.2 Objectives of the Project	2
1.3 Implementation principles	2
1.4 Context and Scope of the Project	4
Context	4
Scope	4
1.5 Participating organisations	5
1.6 Funding arrangements	5
1.7 Management and implementation arrangements	5
1.8 Program Size, Focus and relationship to NCRIS Activities	5
<b>2. PROJECT INFRASTRUCTURE</b>	<b>7</b>
2.1 Infrastructure	7
2.1.1. Data Capture Infrastructure	7
2.1.2. Research Metadata Store Infrastructure	7
2.1.3. Public Sector Data Access Infrastructure	8
2.1.4. Australian Research Data Commons Core Infrastructure	8
2.1.5. Australian Research Data Commons Applications Infrastructure	9
2.2 Early Activity	10
2.2.1 Data Capture Infrastructure	10
2.2.2 Research Metadata Store Infrastructure	10
2.2.3 Public Sector Data Access Infrastructure	10
2.2.4 Australian Research Data Commons Core Infrastructure	11
2.2.5 Australian Research Data Commons Applications Infrastructure	11
2.3 Selection of Further Activity	11
2.3.1 Data Capture Infrastructure	11
2.3.2 Research Metadata Store Infrastructure	12
2.3.3 Public Sector Data Access Infrastructure	12
2.3.4 Australian Research Data Commons Core Infrastructure	12
2.3.5 Australian Research Data Commons Applications Infrastructure	12
2.4 Equipment (if applicable)	12
2.5 Quarterly Milestones	12
2.6 Other Milestones	14
2.7 Project resources	14
<b>3. RISK MANAGEMENT</b>	<b>14</b>
<b>4. ACCESS AND PRICING</b>	<b>17</b>
4.1 General Principles	17
4.2 Information Management and Access	18
4.2.1 Information management	18
4.2.2 Intellectual Property	18
4.3 Access Charges	18
<b>5. OWNERSHIP AND MANAGEMENT</b>	<b>19</b>

5.1 Relation of the Australian Research Data Commons to the Research Community .....	19
5.2 Governance Framework .....	19
5.3 Specific Roles .....	19
5.3.1 Monash University (Lead Institution) .....	19
5.3.2 ANDS Office .....	19
5.3.3 Steering Committee .....	20
5.3.4 Management structure .....	20
5.4 Project Goals and Evaluation.....	20
5.4.1 Goals.....	20
5.4.2 Evaluation.....	21
<b>6. IMPLEMENTATION STRATEGY AND FINANCIAL INFORMATION .....</b>	<b>21</b>
6.1 Implementation Strategy.....	21
6.1.1 Initial Arrangements.....	21
Final Australian Research Data Commons Project Plan .....	22
6.1.2 Implementation from 2010-2011.....	22
6.2 Financial Projections.....	22
6.2.1 Financial Management.....	23
6.2.2 Source of Co-Investment and nature of expenditure .....	23

ANNEX A1 – Project Infrastructure

# 1. PROJECT OVERVIEW

## 1.0 Background and Context

The Australian National Data Service (ANDS) is an initiative of the Australian Government begun under the National Collaborative Research Infrastructure Strategy (NCRIS). Since its development beginning in 2007 and its formal commencement in 2009, it has executed plans to develop Australia's research data infrastructure, capture descriptions of Australia's research data, and to build Australia's research data management capability. The scale of the funding available under NCRIS constrained ANDS to a program of *seeding* Australia's Research Data Commons.

The establishment of the Australian Research Data Commons under the Super Science Initiative will occur with support from the Education Investment Fund (EIF). It will be executed by Monash University through the ANDS office with the support of the current members of the ANDS Collaboration Agreement, namely the Australian national University and the CSIRO.

The Australian Research Data Commons supports the discovery of, and access to, research data held in Australian universities, publicly funded research agencies, government organisations and the Australian Research Collaboration Services (ARCS) Data Fabric for the use of research. This investment will enable the construction of a range of ICT utilities to capitalise on and ensure greater use and re-use of existing data resources, as well as better management of new data generated in Australian research.

Specifically the Project will

- establish the core infrastructure of the Australian Research Data Commons (ARDC)
- construct and install systems that identify and describe significant data holdings for the ARDC
- construct and install systems that permit researchers to explore Australia's research data holdings, and
- construct and install systems that permit researchers to re-use research data through the ARDC.

This Final Australian Research Data Commons Project Plan documents the Project's intended outcomes for the life of the project with milestones for the 2009/10 financial year.

It is acknowledged that the two years of the Australian Research Data Commons Project Plan will coincide with the final two years of the NCRIS Program for the Australian National Data Service and that, while the two activities will be under the same management and be able to maximise opportunities for cost savings, separate reports will be provided to DIISR for the NCRIS service and EIF project.

### 1.1 Definition

The 'commons' is a term that traditionally refers to resources that are made available for community use. ANDS is creating a digital realisation of this concept for research data to enable greater data sharing and re-use.

The Australian Research Data Commons 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.

## 1.2 Objectives of the Project

The objectives of the Project are to:

- establish the ARDC infrastructure and to create the ARDC;
- install infrastructure across the sector so that research data is routinely captured from instruments and devices and published into the ARDC; and
- install infrastructure as needed to ensure that public sector data is visible and accessible through the ARDC.

The ARDC infrastructure will enable:

- any Australian researcher to publish their data through the ARDC; and
- any Australian researcher to discover and exploit Australia's published research data.

The construction of the ARDC will be managed in accordance with the principles set out below:

- (i) Australia's investment in research infrastructure should be planned and developed with the aim of maximising the contributions of the Research and Development (R&D) system to economic development, national security, social wellbeing and environmental sustainability;
- (ii) Infrastructure resources should be focused in areas where Australia is, or has the potential to be, world-class (in both discovery and application driven research) and provide international leadership;
- (iii) Major infrastructure should be developed on a collaborative, national, non-exclusive basis. Infrastructure should serve the research and innovation system broadly, not just the host/funded institutions. Funding and eligibility rules should encourage collaboration and co-investment. It should not be the function of this funding arrangement to support institutional level (or even small-scale collaborative) infrastructure;
- (iv) Access is a critical issue in the drive to optimising Australia's research infrastructure. There should be as few barriers as possible to accessing major infrastructure for those undertaking meritorious research;
- (v) Due regard be given to the whole-of-life costs of major infrastructure, with funding available for operational costs where appropriate, but not from this project; and
- (vi) The Project should seek to enable the fuller participation of Australian researchers in the international research system.

Expected benefits from this project include:

- research data will be routinely published, enhancing the reputation of Australian researchers;
- collaborative research data initiatives of national significance will be facilitated;
- it will be easier for international researchers to work with Australian researchers because of the excellence 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.

## 1.3 Implementation principles

ANDS will establish, operate and provide access to the Project infrastructure and data in accordance with the EIF principles, listed below.

*EIF Principles*

- Principle 1: Projects should address national infrastructure priorities;
- Principle 2: Projects should demonstrate high benefits and effective use of resources;
- Principle 3: Projects should efficiently address infrastructure needs; and
- Principle 4: Projects should demonstrate they achieve established standards in implementation and management.

ANDS will establish, operate and/or provide access to the Australian Research Data Commons and associated services in a manner which:

- takes into account the long-term strategic requirements of relevant research disciplines;
- enhances national and international collaboration in research;
- enhances research capability for relevant disciplines;
- provides for merit-based access to the facilities in accordance with the norms and expectations of the research community;
- focuses on service provision to the research community; and
- focuses on effective use and management of data resources.

In responding to the objectives and program requirements, the Project will continue to follow the established ANDS principles:

- **Commons Framework**  
ANDS has been established 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 Australian Research Data Commons vision, and who provide the most strategic return to the Australian Research Data Commons for the effort expended.
- **Content**  
ANDS is initially focussing on content recruitment into data stores and federation across data stores so as to achieve a wide coverage of data quickly at an agreed level of quality; in future years, further investment may be needed to enhance quality improvement.
- **Service Provision**  
ANDS is focussed on the provision of technology and services, not research and exploration; its programs will develop, integrate, and continually improve production-level systems in support of well-defined services.
- **Strategic Partners**  
ANDS recognises the need to actively engage with institutions relevant to the Australian Research Data Commons, including the Australian Access Federation (AAF) and the Australian Research Collaboration Service (ARCS).
- **Stores**

ANDS assumes an environment where storage and long-term curation occur in nationally and institutionally-supported stores, either existing or brought into being over the life of ANDS. ANDS will facilitate public/restricted access and re-use across these institutional stores. These stores will preferably hold objects described by various discipline-specific and documented metadata schemas. ANDS will work with whatever repositories exist, institutional or discipline-based.

- Sustainability  
Research data management requires a long-term commitment. ANDS has developed its two-year plan 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.

## 1.4 Context and Scope of the Project

### Context

The Australian Research Data Commons 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. Through EIF Funding, 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 the Australian Research Collaboration Service (ARCS) as they establish a national data storage fabric, so that every researcher will be able to store, manage, share and publish their data. ANDS will also work closely with other NCRIS and EIF-funded capabilities, especially expanding existing relationships with the Atlas of Living Australia, Terrestrial Ecosystem Research Network, the Integrated Marine Observation Network, and the Australian Biosecurity Intelligence Network.

### Scope

In the Project, ANDS will create the Australian Research Data Commons that provides a single point to discover descriptions of significant data collections whether they are from an individual research project, a whole research area, significant national collections or from Government data collections. Australian researchers will be able to publish their data, enable them to get credit for their data, share their data, and integrate it with others. The aim of this infrastructure is that Australian researchers will be desirable partners for collaboration because of the ease of data sharing enabled by the Project.

The Project will create the infrastructure that enables:

- capture of both data and metadata that supports reuse;
- the storage of this information through institutional data and metadata stores;
- capture of data collection descriptions for data used by researchers but created by others, in particular Government Departments and National Collections holders;
- the creation of the Australian Research Data Commons that describes research collections, connections between them, and enables the discovery of this information; and
- the tools that enable researchers to exploit this data.

Some of this infrastructure will be established as national services, some will be installed at institutions, and some will be capability specific systems that capture particular forms of data or metadata, or enable discipline specific exploitation of content in the Australian Research Data Commons. The ARDC will have a particular focus on making institutionally supported and nationally significant research data collections

visible through the ARDC. These collections might be created by research institutions or by agencies creating collections relevant to researchers. All publicly funded data collections are within the scope of the ARDC project.

The ARDC links to research data collections, but does not store or manage these collections. It has no brief to be held accountable for these collections, as this resides with the collection holders.

The ANDS Offices will coordinate and manage the creation of and commissioning of the core Australian Research Data Commons services created through the Project. A range of institutions will operate the data capture services, and will be further specified in the Australian Research Data Commons Business Plans.

## 1.5 Participating organisations

It is intended that the original parties to ANDS will be partners in executing the Project. Further partners to ANDS are being considered as a result of this project and are subject to current discussion.

## 1.6 Funding arrangements

The Project will be resourced through the following funding arrangements:

- The Funding will be paid to Monash University, where the ANDS Office will manage the Funds in accordance with this agreement.

Other cash and in-kind contributions will be managed by the ANDS Office.

## 1.7 Management and implementation arrangements

Monash University, through the ANDS Office, will have overall responsibility for the management and implementation of the Project in accordance with the reporting and accountability requirements outlined in the DIISR-Monash agreement. Further details on the nature and governance role of ANDS can be found at Section 5.

The ANDS Office will manage the Project and will sub-contract or partner with organisations to establish, operate and/or provide access to project infrastructure. These subcontractors and partners must be approved by DIISR before the subcontractors' work may commence.

The ANDS Office's sub-contracts will, where appropriate in accordance with the contract, require the sub-contractor to act consistently with the principles set out in this agreement.

## 1.8 Program Size, Focus and relationship to NCRIS Activities

There are five programs proposed to create the infrastructure needed for the ARDC. These complement a revised set of programs in the ANDS NCRIS funded activity that enables greater benefit and synergy. The total set of programs therefore comprise:

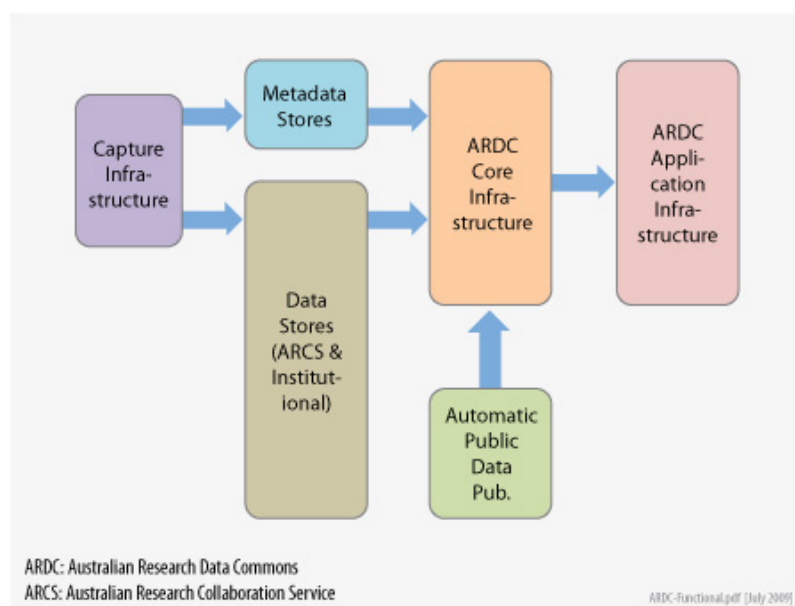
1. **Frameworks and Capabilities** – an ANDS driven program ensuring that institutions have the capability and the research system have the structures in place to enable researchers to manage, publish, share and re-use research data (NCRIS-funded)

2. **Data Capture** – an institutionally based program to automate the capture of data and metadata from instruments in data intensive research
3. **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)
4. **Public Data Access** – an outsourced program of making more public data collections visible and available through the ARDC
5. **Metadata storage** – an institutionally based program that enables metadata to be stored coherently across an institution that supports data management, publishing, sharing and re-use.
6. **ARDC Core** – an ANDS driven program that puts in place the national services that enable research data to be published and discovered.
7. **ARDC Applications** – an outsourced program that establishes national services that enables the exploitation of the data in the ARDC

The next table shows the intended size and focus of the programs.

Programs	EIF (\$)	NCRIS (\$)	%AGE	Focus
Frameworks and Capabilities		4	6.0%	ANDS
Metadata Storage	8		12.0%	Institutions
Data Capture	17		25.5%	Institutions
Seeding the Commons		12	18.0%	Institutions
Public Data Access	10		15.0%	Contractors
ARDC Core	7		10.5%	ANDS
ARDC Applications	5		7.7%	Contractors
EIF Project Management	1		1.5%	ANDS
Project Office		2	3%	ANDS
<b>Total (\$66M)</b>	<b>48</b>	<b>18</b>	<b>100%</b>	

The next diagram shows how the ARDC programs fit together around the creation of the ARDC Core. These five categories of infrastructure address the vision of the Data Commons (as described in *Towards the Australian Data Commons*<sup>1</sup>).



<sup>1</sup> <http://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf>

## 2. PROJECT INFRASTRUCTURE

The infrastructure that is created or purchased will enable the creation of the Australian Research Data Commons, the systems to populate the Commons, and the tools that enable the Commons to be exploited by researchers. ANDS through its technology evaluation criteria specifically seeks to exploit international standards and technologies whenever they exist and are appropriate.

### 2.1 Infrastructure

The amounts paid to Monash University in accordance with the DIISR funding agreement will be used to create and develop the ARDC infrastructure through the five programs in section 1.8 of this project plan. In each category the approach will be to undertake some targeted initial engagements in order to validate the optimal way to build and provide the required infrastructure. There will then be a more structured selection process to commission, develop and deploy the required infrastructure.

#### 2.1.1. Data Capture Infrastructure

The ARDC project will adapt 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. This integration will make it easier for researchers to contribute data to the Data Commons 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 provide 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.

Metadata that describes the data, the experiment, and the wider context will be captured. International standards for metadata and collections descriptions will be strongly preferred. A good example is the UK Science and Technology Facilities Core Scientific Metadata Model (CSMD), as well as crosswalks to other relevant schemata. The use of the SWORD protocol for deposit will be encouraged.

The selection process for this strand of activity will involve working with institutions to identify researchers who are high-profile (Federation Fellow, QE2 holders, etc), working in a data-intensive mode, whose research is dependent on instruments and who are willing to share their resulting data. The number of researchers assisted at each institution will be based on a combination of the institution's level of research intensiveness using an agreed measure and the number of institutions who wish to engage in this program.

#### 2.1.2. Research Metadata Store Infrastructure

The Research Metadata Store Infrastructure consists of software to enable the creation, management, and harvesting of a wide range of metadata about data collections and data objects. This includes rich collections description metadata, including links to the data objects (stored elsewhere, such as on the ARCS Data Fabric), data-level descriptions in order to effectively re-use data (scales, field names,

variables, calibration settings), and study-level descriptions in order to decide whether to re-use the data or not. This program will develop, configure and make available this metadata infrastructure at research producing institutions. Institutions with existing suitable software can have this augmented to the required standard if they choose to. Researchers will be able to automatically or by deliberate choice publish their data using this infrastructure combined with the national services.

The resulting infrastructure components will enhance and complement a diversity of existing data and metadata management solutions that are installed at institutions that produce research data inputs and outputs. ANDS will work with the vendors/creators of the most widely adopted software solutions to ensure support for the range of data-level metadata needed to support ANDS' reusability goals. These include conventional institutional repositories (based on solutions such as VITAL/Fedora, DSpace, or ePrints and usually installed to manage research publication outputs), institutional data stores (often based on SRB/iRODS, or hierarchical storage management solutions), other data/metadata management offerings (such as Arcitecta's MediaFlux), learning object repositories, and document or content management systems. ANDS will also commission an add-on system to provide a complementary metadata store that can be linked to the data store for institutions that would prefer this approach. ANDS will work with vendors/creators to include support for collection objects (and their associated metadata) that meets ANDS' requirements for bindings between collection descriptions, data objects and study-level descriptions. Finally, ANDS will commission the development of software to generate transportable, implementation-independent rich collection descriptions that can be used to enhance discovery services.

The selection processes for this strand of activity will take two forms. In the case of the development of the software, ANDS will choose development groups with the right skills, available capacity, and a demonstrated ability to deliver. This is because of the long lead times for software development. In the case of the specification and deployment of the software, ANDS will select institutions that are ready, are already engaged with ANDS, are broadly representative of the diversity across the sector, have the required level of capability, and who are prepared to commit their own resources.

### **2.1.3. Public Sector Data Access Infrastructure**

This is the infrastructure needed to provide feeds of data collection descriptions from a range of public sector agencies. These will include producers of research data, such as the Bureau of Meteorology, the Australian Bureau of Statistics, GeoScience Australia, and Departments of Primary Industry. It will also include owners of data gathering activities and collections, such as Departments of Primary Industry, museums and libraries, that might be possible inputs to other research activities. The key result is to make existing public sector data resources more discoverable to the research community and work with government agencies to improve access to currently inaccessible data.

The resulting infrastructure components will vary across agencies according to their existing infrastructure and the types of data being made available. In the spatial data domain, ANDS will commission the configuration and deployment of software being developed in the Spatial Information Services Stack project (funded by NeAT) across a range of government agencies. In all cases there will be a strong preference to have data services exposed using relevant international standards such as OGC. ANDS will also commission software components (yet to be specified) to enable the registration of particular datasets at CSIRO, the ABS, and the Bureau of Meteorology.

### **2.1.4. Australian Research Data Commons Core Infrastructure**

This is the infrastructure needed to identify and link Australia's research data now and over time using the ARDC. Consequently the key infrastructure components required are:

- A persistent identifier service for data collections and other digital objects, built on top of the global Handles infrastructure provided by CNRI
- A Collections Registry that stores information about data collections, parties (people and organisations), activities (research programs) and services
- Connections to and links between authoritative sources of information about people, organisations, research activities, funded programs with feeds into the Collections Registry
- Vocabulary services to ensure that the data is of high quality and thus the discovery services produce the most effective experience
- The Research Data Australia webpages about data collections, parties, activities and services that are made available for harvesting by Google

ANDS will establish core Data Commons information infrastructure. Some of this new infrastructure will be developed and operated by ANDS itself and some of it by partners in the sector or in government agencies.

The ANDS persistent identifier service is built on top of the global Handles infrastructure provided by CNRI. ANDS is building extensions to this software to allow for identifier self-minting and management, as well as integrating this capability in other ANDS offerings. ANDS will also work with the German Technische Informationsbibliothek (TIB) to add the ability to register and manage Digital Object Identifiers (DOIs). Because researchers have a number of possible identities, some of them international in scope (e.g. Scopus ID), ANDS will lead activity on International Interoperable Identifier Infrastructure, working directly and also co-ordinating activity in the US, UK and Europe.

ANDS maintains a Collections Registry that stores information about data collections, parties (people and organisations), activities (research programs) and services. This registry is based on the ISO 2146 information model, and is also the target of harvesting by overseas registries (IESR in the UK and Ockham in the US).

In order for ISO 2146 to operate most effectively, it needs to rely on feeds from authoritative sources of information about people, organisations, research activities, and funded programs. ANDS will work with a range of organisations to implement and exploit these feeds. ANDS will also put in place infrastructure to create and manage vocabulary services to be used when selecting metadata elements.

The Collections Registry is used to generate a range of discovery services. The initial offering is Research Data Australia. This is a set of richly interlinked pages based on ISO 2146 entities that are made available for harvesting by Google and other web search engines. This enables discovery of data collections in context. This is intended to be broad discovery rather than subject specific discovery, often better supported by portals such as that being developed by IMOS. Over the next year ANDS will enrich this discovery offering with spatial search and faceted search/browsing, as well as tighter bi-directional integration with the NLA discovery offerings.

#### **2.1.5. Australian Research Data Commons Applications Infrastructure**

Infrastructure needed to fully exploit the data accessible through the Data Commons will be commissioned and installed. This includes infrastructure to support:

- data integration, fusion, and merging,
- data visualisation, and
- data analysis.

The approach taken will be to work closely with leading research groups and Super-Science initiatives on specific tools for the use and re-use of the newly created pool of data assets in the Data Commons.

The resulting infrastructure components will depend on the demands of the specific disciplines being targeted, and the data types involved. However this program targets applications that exploit the ARDC as a whole, enabling researchers in one discipline to benefit from data in another discipline. The overall approach will be based on the exploitation of metadata that are being captured by the ARDC metadata and vocabulary infrastructure. Rather than designing an architectural approach for particular sets of solutions, the ARDC project will deliberately first tackle solutions to specific sets of research problems and then generalise.

## **2.2 Early Activity**

One of the requirements of this project was that ANDS would provide to DIISR an initial specification for the ARDC that would detail \$10M early expenditure to support the creation of the ARDC. This required commitment of funds by September 30<sup>th</sup> 2009. Consequently ANDS produced a description of engagements based on discussions already underway and relationships that had already been established. These descriptions are derived from the initial specification agreed with DIISR prior to June 30<sup>th</sup>, 2009. and updated in this project plan

### **2.2.1 Data Capture Infrastructure**

ANDS will engage with the Australian Synchrotron to determine an initial set of users and beamlines, and commission software to ensure that the image data is captured along with metadata that enables easy transfer to institutional stores, ARCS stores, and institutional metadata stores and easy sharing and publication at appropriate times.

ANDS will commission the existing developers of the FieldHelper tool at the University of Sydney (or an equivalent activity at the University of Southern Queensland) to extend it, adding new functionality and providing integration with the ARCS Data Fabric. This will provide one generic data management solution for the humanities and social sciences.

ANDS will select an initial set of candidate researchers at several research-intensive universities, and undertake six accelerated engagements with these candidates to test the validity of the planned approach. The selection will be based on a number of criteria: is the researcher high-profile, is their research data-intensive (either input or output), and are they prepared to share their outputs?.

ANDS will commission CSIRO to develop tools for the ingestion of high priority data including data from water sensors and data from the Australia Telescope National Facility.

### **2.2.2 Research Metadata Store Infrastructure**

ANDS will partner with the Australian Digital Futures Institute at USQ (subject to a suitable agreement being reached) to develop a solution that meets the requirements listed above. This solution would be integrated with other ANDS services and the ARCS Data Fabric.

ANDS will also commission software for early installation at Monash University and the Australian Synchrotron to improve metadata management for data objects.

### **2.2.3 Public Sector Data Access Infrastructure**

ANDS will commission a fast initial rollout of the SISS offering to an initial group of institutions (that includes several Departments of Primary Industry with whom AuScope is already engaged), who hold well-managed spatial data and are willing to share it.

ANDS will work with the CSIRO to identify and expose for discovery an initial set of data collections .

ANDS will also fund the development of the National Criminal Justice Data Network to aggregate and provide increased access to Criminal Justice data from a range of state-based data holders in this area, including NSW Bureau of Crime Statistics and Research, NSW Dept of the Attorney-General, the SA Office of Crime Statistics and Research, SA Dept of the Attorney-General, the NT Department of Justice, the Australian Institute of Criminology, and the ABS National Centre for Crime and Justice Statistics.

#### **2.2.4 Australian Research Data Commons Core Infrastructure**

ANDS will extend its persistent identifiers offering, and implement extensions to the existing software used to generate the Research Data Australia pages. A key requirement for discovery and linking between data sources is effective data contributor identification. ANDS will fund the NLA to expand their People Australia offering, integrating the NICNAMES project at Swinburne University, to provide a single authoritative source of author identifiers for Australia. ANDS will fund the University of Melbourne and Monash University to develop infrastructure to ensure that the ANDS Collections Registry receives accurate and timely feeds of information about all research parties and activities.

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

In the area of data integration/fusion/merging, the ARDC project will fund AuScope through CSIRO to build new spatial data services out of existing services. The project will provide the Atlas of Living Australia with the ability to use annotation to improve data quality; and to integrate a variety of literary data sources for improved discovery. Lastly, the project will integrate a large number of sources of data associated with the coast from many disciplines to enable exploration of the data through common metadata and discovery tools.

In the area of data visualisation, the ARDC project will commission a tool that enables the visualisation of the connections between projects, researchers, and their research data to show the relationships between them.

In the area of data analysis, the ARDC project will initiate an activity with the University of Queensland as candidate service developer and a service provider (to be identified) to mark up existing data with a Who/What/When/Where data combination service and demonstrate new analysis opportunities on top of this marked-up data.

All of these activities have target owners and operators of the resulting services.

### **2.3 Selection of Further Activity**

Whilst very substantial, the EIF ARDC program will not completely solve research data issues in Australia, and consequently there is a need to choose where effort will be applied. Sometimes this will be to determine what infrastructure is produced, and sometimes where it is deployed. As was indicated in the table in section 1.7, the locus of these decisions varies by programs – sometimes ANDS will decide based on an agreed roadmap, and sometimes it will be an institutional decision using ANDS guidelines.

#### **2.3.1 Data Capture Infrastructure**

There are more research projects with high data capture needs than can be supported by the funds available for the data capture program. Consequently choices must be made. Rather than using a merit allocation approach, ANDS believes that it is appropriate to engage in a negotiation with research institutions on a joint selection of projects based on the institutional priorities and ANDS priorities of data intensity, ability to capture data and metadata automatically, research quality, the coverage of relevant research data, and disposition to share data. The level of activity would be based on the level of research activity at the institution.

ANDS will allocate some of the funds for this program to nationally significant research data initiatives that are supported by several institutions. For example, if 3 institutions selected activities that were related to a national initiative, ANDS would further augment the available funds to achieve a nationally significant outcome.

### **2.3.2 Research Metadata Store Infrastructure**

Every research institution that wishes to engage with ANDS will benefit from a coordinated approach to the management and storage of the metadata associated with its research data. Consequently this program needs to support any organisation that wishes to implement a solution. However, the program cannot fund a unique solution at every institution, so ANDS will support a small number of approaches that can be widely used. Consequently ANDS will engage in negotiations with institutions around solutions implementation, chosen from a small set of possible approaches that meet their needs.

### **2.3.3 Public Sector Data Access Infrastructure**

The selection of further partnerships to make research data discoverable through the ARDC will be based on the public sector dataholders' interest in making data available, their commitment to ongoing data services beyond the life of the project, and the importance of the data to Australian researchers. Consequently selection will occur through discussion with both the data holders and representatives of research disciplines and capabilities.

### **2.3.4 Australian Research Data Commons Core Infrastructure**

The selection process for this strand of activity will be based on the development of a Technical Roadmap of what is required. Identification of developers for components in the technical roadmap will be a mix of selection by ANDS staff (if there is thought to be only one candidate provider) or an Expression of Interest process otherwise.

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

The selection process for this strand of activity will be identical to that for the ARDC Core Infrastructure program.

## **2.4 Equipment (if applicable)**

The software systems that enable institutional research data capture and storage will need hardware. However this is not of a highly specialised nature so most hardware should be purchased through standard institutional processes.

## **2.5 Quarterly Milestones**

Monash University must complete the following Milestones by the specified completion date:

<b>Milestone Date</b>	<b>Data Capture Infrastructure</b>	<b>Metadata Store Infrastructure</b>	<b>ARDC Access Infrastructure</b>	<b>ARDC Core Infrastructure</b>	<b>ARDC Application Infrastructure</b>
<b>09Q3</b>	Requirements identified	Identify software targets and required changes; Engagement strategy for deployment complete	Engagement strategy complete	Initial utilities launched; Planning for utilities enhancement	Requirements identified
<b>09Q4</b>	EOI called	Commence engagement with software owners; Engagement with data providers on deployment;	Requirements identified (inputs)	Development of enhanced initial utilities; Analysis/specification of first batch of new utilities; load monitoring and projection	EOI called
<b>10Q1</b>	Community development partners identified and contracted; Software specifications complete	First batch of data management software changes available for deployment;	Software commissioned (inputs)	Testing and release of enhanced initial utilities; Development of first batch of new utilities	Community development partners identified and contracted; Software specifications complete
<b>10Q2</b>	Software development commences	Second batch of data management software changes available for deployment; First batch deployments complete	Compliance testing (inputs); Requirements identified (outputs)	Testing and release of first batch of new utilities; Production user feedback received; Analysis/specification of second batch of new utilities; Specify hosting environment enhancements;	Software development commences
<b>10Q3</b>	Software testing commences	Third batch of data management software changes available for deployment; Second batch deployments complete	First set of data feeds activated (inputs); Software commissioned (outputs)	Implement upgraded hosting	Software testing commences
<b>10Q4</b>	Initial release of software deployed; Production user feedback received; Design and specs for final release complete	Fourth batch of data management software changes available for deployment; Third batch deployments complete	Second set of data feeds activated (inputs); Compliance testing (outputs)	Development of second batch of new utilities complete	Initial release of software deployed; Production user feedback received; Design and specs for final release complete
<b>11Q1</b>	Final release developed	Fourth batch deployments complete	First set of data feeds activated (outputs)	Testing and release of second batch of new utilities	Final release developed
<b>11Q2</b>	Final release deployed		Second set of data feeds activated (outputs)		

## 2.6 Other Milestones

Milestones past July 2010 will be provided in the Australian Research Data Commons Business Plan, as will confirmed co-investment from ANDS partners and other collaborating parties in this Project.

## 2.7 Project resources

The Project resources are shown in the financial tables in section 6.2

## 3. RISK MANAGEMENT

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

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

### *Political and Governance*

#### **Risk 1 – There are persistent negative perceptions of the Project among funding agencies and influential groups leading to lack of buy-in**

Risk Factors:

- Lack of confidence in governance, management, or Project delivery;
- Perceptions of slow engagement with areas of the sector;
- Change of policy with regard to the policies around publicly funded research data; and
- The Project message and brand are diluted by the many partnerships.

Risk Mitigations:

- 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;
- Provide a central point where progress towards the Australian Research Data Commons can be tracked by metrics such as number of collections available, and numbers of datasets accessed; and
- The Project will clearly articulate its message and brand.

#### **Risk 2 - The Project is not managed effectively**

Risk Factors:

- Lack of effective mechanisms for planning, leadership and management;

- The Project staff see their primary loyalty as being to their employing organisation not the Project;
- Collaboration between the Project and across locations is not effective;
- Insufficient Project staff to manage and coordinate work; and
- State based staff will have mixed allegiances.

Risk Mitigations:

- Management and planning processes that include formal reporting and regular reviews are put in place to ensure the efficient conduct of the Project;
- Regular meetings of Project staff are held to build a team approach;
- Staffing levels are monitored and adjusted as required; and
- Contracts between state based organisations that host Project staff will be developed that ensure that staff are clear about their role.

*Relationships*

**Risk 3 - The Project's external stakeholders are not effectively engaged**

Risk Factors:

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

Risk Mitigations:

- Maximise the effectiveness of connections between the Project and related PFC and other initiatives, including involvement of groups outside ANDS in the ANDS Policy Forum, the ANDS Technical Forum, and the ANDS Content Forum;
- Ensure continuing wide consultation following the consultation on the Draft Final Australian Research Data Commons Project Plan;
- 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; and
- Performance measurement for the Project should include effective stakeholder engagement.

**Risk 4 – The Project's service providers do not contribute effectively to the Project**

Risk Factors:

- Lack of effective arrangements in place to ensure the required utilities services are provided to an agreed service level; and
- Service providers see themselves as disconnected from the Project's decision-making or strategic planning.

## Risk Mitigations:

- Formal procurement processes to ensure that the requirements are understood and that potential suppliers meet the set criteria;
- Ongoing contract management to ensure the continuation of required services to the contracted service levels;
- Effective vendor management approaches are put in place; and
- Involvement of service providers in planning sessions.

*Impact***Risk 5 – Data providers/federators do not make their data available through the Project**

## Risk Factors:

- 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; and
- Existing data federations see insufficient value in making their data available.

## Risk Mitigations:

- 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;
- Provide targeted assistance to data federations to assist with integration into the ARDC; and
- Recommend that funding be linked to the provision of data via the ARDC as it becomes available.

**Risk 6 – Re-users of research data do not use ARDC to discover and access it**

## 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; and
- 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; and
- ensure that it is easy to re-purpose ARDC accessible data.

*Resourcing***Risk 7 – High quality Project Staff are hard to recruit**

Risk Factors:

- Limited availability of skilled staff to perform roles in the Project;
- Limited tenure roles potentially on offer within the Project are not attractive to candidates; and
- People with second order skills may end up being employed because of staff shortages.

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; and
- Investigate non-traditional sources of potential staff.

**Risk 8 – Funding for the Project is inadequate to achieve its objectives**

Risk Factors:

- Federal budget constraints and/or competition from other programs results in reduced funding;
- The Project fails to manage expectations of what can be achieved within allocated budget; and
- Aspects of the Project can't be realised below a certain funding threshold.

Risk Mitigations:

- Ensure that the Project delivers early wins to generate momentum;
- Use robust and transparent selection processes for future activity;
- Manage expectations through frequent communication via a range of channels; and
- focus the Project to improve likelihood of successful outcomes.

**4. ACCESS AND PRICING****4.1 General Principles**

The mechanisms for deciding access and pricing to the project infrastructure will be consistent with ANDS services. However, as the services generated as outputs by each program are significantly different, the access regimes are described on a program basis.

Generally speaking, the Project, through 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 for publicly funded researchers.

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 Australian Research Data Commons and therefore supported by ANDS utilities and other support

activities. The establishment of the Australian Access Federation may facilitate the provision of appropriate access regimes.

## **4.2 Information Management and Access**

### **4.2.1 Information management**

Information management, including intellectual property and access charging issues, will be managed within the existing ANDS arrangements.

ANDS will document Project processes, decisions and deliverables, and keep records appropriate to enable scrutiny of its decisions and outcomes.

### **4.2.2 Intellectual Property**

For data and data products produced or created under the Project under the auspices of ANDS, it is the intent of ANDS that these be openly available and provided unencumbered. Any Intellectual Property Rights (IPR) associated with raw data are subject to the policies of the institution that produced the data. Any IPR associated with ANDS value-added products produced under the auspices of ANDS lie with Monash University. The IPR owning institutions will issue a non-exclusive licence to any user of the data or value added products.

Any IPR existing at the commencement date of ANDS will be recognised and respected and will only be used as authorised by the owner of the IPR or as permitted under Law.

To minimise conflicts over IPR arrangements, all data, products and services developed under the auspices of ANDS will be free of encumbrances (i.e. users should be able to use and apply the data in whatever way they see fit without restriction other than that set out below).

In some instances several parties will work collaboratively to develop value-added products or services. In these instances, except in the circumstances where ANDS funds development of data products and/or services, or where subject to another project agreement, IPR will be equally shared amongst the parties as tenants in common, regardless of the individual level of contribution made by a Project Participant in developing the material.

Because the ARDC provides links to data collections that are held and managed by their institutional owners, ANDS will only hold publicly available collections descriptions. It will have no IPR over the collections and no liability for those collections.

## **4.3 Access Charges**

The descriptions of data collections from the nationally supported infrastructure will be accessible to all users through Australia Research Data Commons. Ordinarily, automated electronic access will be provided free of charge. ANDS will maintain the infrastructure and underwrite the ongoing data flow for the automated delivery system.

## 5. OWNERSHIP AND MANAGEMENT

### 5.1 Relation of the Australian Research Data Commons to the Research Community

Delivering the Project outcomes and translating them into enduring changes will depend on the contribution, cooperation and cohesion of many higher education and research institutions as well as the major government agencies from which researchers access primary information. The Project recognises that activity, responsibility, and change must happen well beyond a small circle of partners. Initially these are Monash University, the Australian National University (ANU) and the Commonwealth Scientific Industrial and Research Organisation (CSIRO), who currently share direct responsibilities for ANDS governance and management. The Project will focus strongly on engagement with the whole sector, sourcing expertise for projects, activities and services from wherever that expertise resides.

The Project will carefully structure its governance to ensure broad representation from the sector. Decisions will be made in as open a way as possible, consistent with the acceptance and management of risk by the lead agency. In order to learn from others and build a consensus around what needs to be done, the Project will also engage with its various stakeholders in a number of ways.

### 5.2 Governance Framework

In establishing the Governance and Management arrangements for the Project, Monash will seek to continue those employed by ANDS but with wider participation. Those arrangements, previously described in the collaboration agreement as accepted by DIISR, will be described in a separate Collaboration Agreement once negotiations are able to be concluded with the existing ANDS partners, and then with potential additional ANDS partners. 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.

Particular elements that may be of interest are detailed below including features of the ANDS Governance framework to the extent that they relate to the Project.

### 5.3 Specific Roles

#### 5.3.1 Monash University (Lead Institution)

Monash University has entered into an agreement with DIISR for implementation of the Project, receive EIF Funds and will be accountable to DIISR for execution and performance of Project. This will continue existing and parallel arrangements between Monash and DIISR in relation to NCRIS funding currently received by Monash for the implementation of the NCRIS ANDS agreement.

Monash hosts and operates one of the ANDS Offices, which will be used to manage the Project. Monash has appointed the independent Chair of the Steering Committee after consultation with the ANDS partners and formally includes the independent Chair in the performance management arrangements of the Executive Director of ANDS.

#### 5.3.2 ANDS Office

ANDS and Project staff will work collaboratively with each other and support activities across ANDS and the Project. Some will be located at ANDS Member institutions and others '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.

Project 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.

The Executive Director of ANDS is Dr Ross Wilkinson.

### 5.3.3 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 June 2009 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 Lee Taylor (NSW Department of Health)
- Executive Director (ex-officio): Dr Ross Wilkinson (Australian National Data Service)

It is anticipated that the ANDS Steering Committee membership will be expanded to incorporate the additional requirements of the Project.

### 5.3.4 Management structure

ANDS is currently managed by a full time executive staff comprising an Executive Director (located at Monash), and three Deputy Directors (2 at Monash, 1 at ANU) as currently agreed under the ANDS Collaboration Agreement.

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

## 5.4 Project Goals and Evaluation

### 5.4.1 Goals

The primary aim of ANDS is to enable Australian researchers to re-use and share research data more often. The scope of the project is set out in section 1.4.

The implementation of the Australian Research Data Commons will enable researchers to routinely manage their own research data better, make it easier to share with partners, easier to publish, and easier for other researchers to find relevant data to solve new problems.

The outcomes for the National Innovation System are intended to be more efficient research, by reducing reproduction of data, more effective research due to larger problems being tackled collaboratively, and more cited research output as research data is published with recognised benefits to research paper publication.

#### **5.4.2 Evaluation**

The aim of the Project is that researchers reuse research data more often, that they benefit from other researchers' investigations and that they are able to answer new research questions as a result of better, wider, and easier data access. Because these aims are hard to measure early in the life of this Project, the Project should instead measure:

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 that are building or have installed Project infrastructure;
3. The number of times a search is initiated with an Australian Research Data Commons discovery service;
4. The number of times an Australian Research Data Commons data page is accessed, and
5. The satisfaction of researchers and partners with ANDS services as measured by an annual survey; There are two measures that the Project will not have full control over, but that are important and will measure the Project's success in influencing others' behaviour:
6. The number of research data sets in harvestable repositories; and
7. The number of research data sets with persistent identifiers.

These indicators will be measured with a combination of qualitative and quantitative measures. It will also be important that the information obtained for the evaluation be used specifically to judge the prospect and extent of the project becoming sustainable beyond the term of funding.

## **6. IMPLEMENTATION STRATEGY AND FINANCIAL INFORMATION**

### **6.1 Implementation Strategy**

#### **6.1.1 Initial Arrangements**

Core tasks to be completed by Monash University between 10 June 2009 and 30 September 2009 were:

1. A preliminary specification for investments towards establishing the Australian Research Data Commons was provided to DIISR on June 24<sup>th</sup>, 2009. This document detailed:
  - arrangements for the provision of the initial infrastructure elements in the first 12 months of the Project; and
  - arrangements for the expenditure of the first payment under this agreement.
2. A draft of the Final Australian Research Data Commons Project Plan addressing the following points was provided to DIISR, including:
  - a clear characterisation of the components of the national infrastructure to be established;

- priority recommendations for the total investments to be undertaken, based on the consultations undertaken to date;
  - an outline activity plan that would deliver the desired outcomes; and
  - a high level budget breakdown into those component activities.
3. A consultation process designed to produce a Final Australian Research Data Commons Project Plan was conducted between 21<sup>st</sup> of August and 14<sup>th</sup> of September 2009.
  4. An extension of the current ANDS governance arrangements to include the Project activities was agreed to by the ANDS Steering Committee on the 6<sup>th</sup> of August.
  5. Implement the relevant components of the initial \$10,000,000 plan for 2009/10, as outlined in the Preliminary Specifications.

### Final Australian Research Data Commons Project Plan

The ANDS Office has prepared this document, the Final Australian Research Data Commons Project Plan after being informed by a consultation process with the sector, conducted by DIISR in collaboration with the Australian eResearch Infrastructure Committee (AeRIC), to determine how to meet the requirements of the EIF Funding, based specifically on the proposition that:

“The Australian Research Data Commons will support the discovery of, and access to, research data held in Australian universities, publicly funded research agencies government organisations and the ARCS Data Fabric for the use of research.”

This investment should enable the construction of a range of ICT utilities to capitalise on and ensure greater use and re-use of existing data resources, as well as better management of new data generated in Australian research.”

#### 6.1.2 Implementation from 2010-2011

This Final Australian Research Data Commons Project Plan includes overall planning and specific milestones for the period to 30 June 2011.

### 6.2 Financial Projections

A summary of the Financial Projections for the 2009/2010 and 2010/11 years is noted below:

Facility	Operator	2009/10 Budget	2010/11 Budget
Data Capture Infrastructure	Institutions	\$8,500,000	\$8,500,000
ARDC Access Infrastructure	Institutions	\$5,000,000	\$5,000,000
Metadata Store Infrastructure	CSIRO/Institutions	\$4,000,000	\$4,000,000
ARDC Core Infrastructure	ANU/Monash	\$3,500,000	\$3,500,000
ARDC Applications Infrastructure	Institutions	\$2,500,000	\$2,500,000
ARDC Project Management	Monash	\$500,000	\$500,000
<b>Total for 2009/10</b>		<b>\$24,000,000</b>	<b>\$24,000,000</b>

The breakdown of expenditure by anticipated ANDS staff and external staff is shown below:

Spending 2009-2011	Total Expenditure	Internal staff	Internal Expenditure	External	External Expenditure

				staff	
Data Capture	\$17,000,000	5	\$1,250,000	126	\$15,750,000
Public Data Access	\$10,000,000	5	\$1,250,000	70	\$8,750,000
Metadata Storage	\$8,000,000	2	\$500,000	60	\$7,500,000
ARDC Core	\$7,000,000	16	\$4,000,000	24	\$3,000,000
ARDC Apps	\$5,000,000	2	\$500,000	36	\$4,500,000
Project Management	\$1,000,000	3	\$750,000		\$250,000
Total	\$48,000,000	33	\$8,375,000	316	\$39,625,000

The location of internal expenditure is shown below:

Internal staffing over 2 years		2009-10	2010-11	Total
Monash staff	10	\$1,250,000	\$1,250,000	\$2,500,000
ANU/CSIRO Staff	23	\$2,875,000	\$2,875,000	\$5,750,000
Admin costs		\$125,000	\$125,000	\$250,000
Total		\$4,250,000	\$4,250,000	\$8,500,000

### 6.2.1 Financial Management

The Project is subject to the financial management arrangements at Monash University, and disburses funds on the basis of these arrangements.

### 6.2.2 Source of Co-Investment and nature of expenditure

The nature of the 2 year plan means that contemporaneous co-investment is not sought. Instead ANDS has built on previous investments in data and expressed future intentions by our collaborators.

ANNEX A1 – Description of Project infrastructure agreed to be initiated in 2009/10 in the ARDC Preliminary Specification

<b>Infrastructure</b>	<b>Data Capture</b>	<b>Indicative Expenditure</b>
Data Capture Infrastructure	Install and commission systems to automatically capture data and metadata from <ul style="list-style-type: none"> <li>- the Australian Synchrotron</li> <li>- the Auscope project</li> <li>- the Integrated Marine Observation Service</li> <li>- the Atlas of Living Australia</li> <li>- the Australian Biosecurity Intelligence Network</li> </ul>	\$3M
Public Access Infrastructure	Engage with the Bureau of Meteorology and CSIRO on making their water data available through the ARDC	\$1M
Metadata Store Infrastructure	<ul style="list-style-type: none"> <li>- Create a metadata store solution that connects the data stores, institutions and the ARDC</li> <li>- Install ARDC aware metadata stores at three research institutions</li> </ul>	\$1M
ARDC Core Infrastructure	<ul style="list-style-type: none"> <li>- Install robust persistent identification services</li> <li>- Install and populate a collections registry</li> <li>- Install initial discovery services</li> </ul>	\$2M
ARDC Applications Infrastructure	<ul style="list-style-type: none"> <li>- Build services to annotate existing data</li> <li>- Create tools to integrate marine data sets using international standards</li> <li>- Create tools to assemble geospatial data services</li> </ul>	\$2M
ARDC Project Management	<ul style="list-style-type: none"> <li>- Create and maintain ARDC component specifications and descriptions</li> <li>- Specify and commission software contracts</li> <li>- Oversee projects</li> </ul>	\$1M
<b>Total</b>		<b>\$10M</b>